

Universidade do Minho
Escola de Ciências

Helena Isabel da Rocha Piairo

Biomonitorização Ambiental
em Portugal Continental
Uma Análise Espacial e Temporal

Helena Isabel da Rocha Piairo Biomonitorização Ambiental em Portugal Continental
Uma Análise Espacial e Temporal



Universidade do Minho
Escola de Ciências

Helena Isabel da Rocha Piai

Biomonitorização Ambiental
em Portugal Continental
Uma Análise Espacial e Temporal

Tese de Mestrado
Estatística

Trabalho efetuado sob a orientação da
Doutora Raquel Menezes e Doutora Inês Sousa

Agosto de 2013

Agradecimentos

Os meus agradecimentos são dirigidos a todos aqueles que me ajudaram na realização deste trabalho, nomeadamente:

- às Professoras Doutora Raquel Menezes e Doutora Inês Sousa por todo o apoio e atenção;
- a todos os outros Professores e colegas que me acompanharam;
- aos meus amigos e familiares por todo o apoio e compreensão;
- ao Centro de Biologia Ambiental da Universidade de Lisboa por ceder o conjunto de dados analisados neste trabalho, parceiro do projeto FCT, *"Modelos conjuntos para processos espaço-temporais e respetivo desenho amostral, em Ciências do Ambiente e Saúde"*, PTDC/MAT/112338/2009, no qual este trabalho se enquadra.

Este trabalho foi financiado por Fundos FEDER através do Programa Operacional Fatores de Competitividade - COMPETE e por Fundos Nacionais através da FCT - Fundação para a Ciência e a Tecnologia no âmbito do projeto PEst-C/MAT/UI0013/2011.

Resumo

Determinados organismos vivos, como por exemplo os musgos, apresentam capacidades de acumulação de elementos químicos e uma sensibilidade à poluição e, por esse motivo, têm sido largamente utilizados como forma de quantificar a concentração de poluentes no ambiente, processo denominado por biomonitorização. Esta metodologia é considerada bastante simples e com uma elevada resolução espacial (correspondendo, no caso em estudo, a Portugal continental) e alguma resolução temporal, proporcionando uma interpretação relevante em termos da qualidade ambiental.

Neste estudo pretende-se fazer uma análise geoestatística dos dados recolhidos pelo Centro de Biologia Ambiental da Universidade de Lisboa, através de uma rede de biomonitorização abrangendo Portugal continental, com o objetivo de determinar a deposição de vários elementos químicos tais como As, Cd, Cr, Cu, Fe, Hg, Mn, Ni, Pb, Zn e V. Os resultados do sistema de biomonitorização ambiental permitiram, ainda, a identificação das localizações que apresentam características muito semelhantes na região de Portugal, através da análise de *clusters*, permitindo a sinalização das regiões mais poluídas; bem como o reconhecimento das relações lineares existentes entre os conjuntos múltiplos de dados, através da análise em componentes principais.

No entanto, usando este método de análise multivariada não se tem acesso à identificação das diferentes fontes de contaminação ambiental subjacentes às elevadas concentrações dos elementos químicos nos musgos, constituindo variáveis que não podem ser medidas diretamente no meio ambiente, designadas por variáveis latentes. Desta forma, recorre-se ao método estatístico de análise fatorial, o qual permite que estas variáveis possam ser tratadas como contínuas, resultando da separação de um número limitado de associações discretas, cujo método fornece um meio para testar se esta divisão é estatisticamente significativa. Portanto, definiram-se as concentrações de um dado elemento químico através de uma combinação linear de todas as fontes de poluição e não uma nova variável que resulta da combinação de todos os elementos.

Com o objetivo de obter mapas de valores preditos para as diferentes fontes de poluição, aplicaram-se técnicas de modelação geoestatística sobre os *scores* das variáveis latentes, obtidas através da metodologia referida. Deste modo, consegue-se resumir a informação obtida através das concentrações dos vários elementos químicos nos musgos, determinando-se simultaneamente os locais que padecem de determinada fonte de poluição.

Palavras-chave: Biomonitorização, Qualidade Ambiental, Análise Multivariada, Variáveis Latentes, Geoestatística.

Abstract

Some organisms, such as mosses, have capabilities of accumulation of chemical elements and a sensitivity to pollution, therefore, have been widely used as a way of quantifying the concentration of pollutants in the environment, a process referred as biomonitoring. This methodology is considered quite simple and with a high spatial (corresponding, in this study, to mainland Portugal) and some temporal resolution, which provides a significant interpretation in terms of environmental quality.

With this study, we aim to make a geostatistical analysis of the data collected by the Center of Biology Environmental, University of Lisbon, through a biomonitoring network covering mainland Portugal, aiming to determine the deposition of various chemical elements such as As, Cd, Cr, Cu, Fe, Hg, Mn, Ni, Pb, Zn and V. The results of the environmental biomonitoring system allowed also the identification of locations that have very similar characteristics in the region of Portugal, through the cluster analysis, allowing to identify the most polluted regions; as well as the recognition of linear relationships between multiple data sets, through the principal components analysis.

However, using this method of multivariate analysis we do not have access to the identification of different sources of environmental contamination underlying high concentrations of chemical elements in mosses, constituting variables that can not be measured directly in the environment, referred by latent variables. So, we use factor analysis method, which allows these variables to be treated as continuous, resulting from the separation of a limited number of discrete associations, whose method provides a way to test if this division is statistically significant. Therefore, we define the concentration of a specific chemical element by a linear combination of all sources of pollution and not a new variable which results from the combination of all chemical elements.

In order to obtain maps of predicted values for the different sources of pollution, we applied geostatistical modelling techniques on scores of latent variables, obtained by the mentioned methodology. Thus, it was possible to summarize the information obtained from the concentrations of the various chemical elements in mosses and determine, simultaneously, the locations that suffer from a particular source of pollution.

Keywords: Biomonitoring, Environmental Quality, Multivariate Analysis, Latent Variables, Geostatistics.

Contents

1	Introdução	1
2	Componentes do Sistema Ambiental	4
2.1	Estado da arte	6
2.2	Elementos químicos analisados	8
2.3	Amostragem	11
3	Enquadramento Teórico	13
3.1	Métodos de análise multivariada	13
3.1.1	Análise de <i>clusters</i>	14
3.1.2	Análise em componentes principais	16
3.1.3	Análise fatorial	19
3.1.3.1	Análise fatorial exploratória	19
3.1.3.2	Análise fatorial confirmatória	21
3.2	Métodos de análise geoestatística	23
3.2.1	Hipóteses requeridas	24
3.2.2	Variograma empírico	26
3.2.3	Estimação	28
3.2.4	Validação cruzada	31
3.2.5	Predição	32
4	Biomonitorização da Deposição Atmosférica dos Elementos Químicos	34
4.1	Caracterização dos elementos químicos - análise univariada	34
4.1.1	Análise exploratória dos elementos químicos	34
4.1.2	Análise e modelação da continuidade espacial dos elementos químicos	37

4.2	Reconhecimento dos locais poluídos e das relações subjacentes aos elementos químicos - análise multivariada	49
4.2.1	Análise exploratória dos locais mais poluídos em Portugal - análise de <i>clusters</i> . .	51
4.2.2	Análise exploratória das relações existentes entre os elementos químicos - análise em componentes principais	57
5	Identificação e Modelação Espacial das Principais Fontes de Poluição	63
5.1	O porquê da utilização da análise fatorial	64
5.2	Identificação das principais fontes de poluição	65
5.3	Modelação espacial das principais fontes de poluição	69
6	Conclusão	75
	Bibliografia	77
	Distritos de Portugal continental	80
	Valores das concentrações dos elementos químicos nos musgos	81
	Diagramas de extremos e quartis dos elementos químicos	85
	Análise geoestatística dos elementos químicos	87
	Diagramas de extremos e quartis das fontes de poluição	118
	Análise geoestatística das fontes de poluição	119

Lista de Figuras

2.1	Esquema de um estudo de biomonitorização, adaptado de Figueira (2002).	5
2.2	Mapas de Portugal continental com as diferentes localizações amostradas em cada campanha. .	11
4.1	Variogramas amostrais das concentrações de As nos musgos observadas em cada campanha e respetivos modelos esférico (à esquerda) e exponencial (à direita), estimados pelo método de mínimos quadrados.	40
4.2	Superfícies preditas das concentrações de As nos musgos observadas em cada campanha. . . .	40
4.3	Superfícies preditas das concentrações de Cd nos musgos observadas em cada campanha. . . .	41
4.4	Superfícies preditas das concentrações de Cr nos musgos observadas em cada campanha. . . .	42
4.5	Superfícies preditas das concentrações de Cu nos musgos observadas em cada campanha. . . .	43
4.6	Superfícies preditas das concentrações de Fe nos musgos observadas em cada campanha. . . .	44
4.7	Superfícies preditas das concentrações de Mn nos musgos observadas em cada campanha. . . .	45
4.8	Superfícies preditas das concentrações de Ni nos musgos observadas em cada campanha. . . .	46
4.9	Superfícies preditas das concentrações de Pb nos musgos observadas em cada campanha. . . .	47
4.10	Superfícies preditas das concentrações de Zn nos musgos observadas em cada campanha. . . .	48
4.11	Matrizes das correlações entre as concentrações dos vários elementos químicos nos musgos observadas em cada campanha. Representam-se os histogramas (diagonal) dos elementos químicos, os coeficientes de correlação (acima da diagonal) e os diagramas de dispersão (abaixo da diagonal).	51
4.12	Dendrogramas das concentrações dos vários elementos químicos nos musgos observadas em cada campanha, utilizando o método de agrupamento de <i>Ward</i> e a medida de dissemelhança de <i>Chebychev</i>	52
4.13	Mapas de Portugal continental com as diferentes localizações amostradas em cada campanha subdivididas pelos vários grupos.	53
4.14	Planos dos <i>scores</i> da segunda versus terceira componentes principais determinadas em cada campanha.	62
5.1	Especificação pictográfica dos modelos de análise fatorial das fontes de poluição subjacentes às concentrações dos elementos químicos nos musgos observadas em cada campanha.	68
5.2	Valores dos <i>scores</i> das fontes de poluição subjacentes à campanha de 1992. Os valores são proporcionais à área do círculo.	70

5.3	Superfícies preditas das concentrações das fontes de poluição subjacentes à campanha de 1992.	70
5.4	Valores dos <i>scores</i> das fontes de poluição subjacentes à campanha de 1997. Os valores são proporcionais à área do círculo.	71
5.5	Superfícies preditas das concentrações das fontes de poluição subjacentes à campanha de 1997.	71
5.6	Valores dos <i>scores</i> das fontes de poluição subjacentes à campanha de 2002. Os valores são proporcionais à área do círculo.	72
5.7	Superfícies preditas das concentrações das fontes de poluição subjacentes à campanha de 2002.	72
5.8	Valores dos <i>scores</i> das fontes de poluição subjacentes à campanha de 2006. Os valores são proporcionais à área do círculo.	73
1	Mapa de Portugal continental, referenciando as regiões contidas na NUT II, bem como os distritos de Portugal.	80
2	Valores das concentrações de As nos musgos observadas em cada campanha. Os valores das concentrações são proporcionais à área do círculo.	81
3	Valores das concentrações de Cd nos musgos observadas em cada campanha. Os valores das concentrações são proporcionais à área do círculo.	81
4	Valores das concentrações de Cr nos musgos observadas em cada campanha. Os valores das concentrações são proporcionais à área do círculo.	82
5	Valores das concentrações de Cu nos musgos observadas em cada campanha. Os valores das concentrações são proporcionais à área do círculo.	82
6	Valores das concentrações de Fe nos musgos observadas em cada campanha. Os valores das concentrações são proporcionais à área do círculo.	82
7	Valores das concentrações de Hg nos musgos observadas em cada campanha. Os valores das concentrações são proporcionais à área do círculo.	83
8	Valores das concentrações de Mn nos musgos observadas em cada campanha. Os valores das concentrações são proporcionais à área do círculo.	83
9	Valores das concentrações de Ni nos musgos observadas em cada campanha. Os valores das concentrações são proporcionais à área do círculo.	83
10	Valores das concentrações de Pb nos musgos observadas em cada campanha. Os valores das concentrações são proporcionais à área do círculo.	84
11	Valores das concentrações de Zn nos musgos observadas em cada campanha. Os valores das concentrações são proporcionais à área do círculo.	84
12	Valores das concentrações de V nos musgos observadas em cada campanha. Os valores das concentrações são proporcionais à área do círculo.	84
13	Diagramas de extremos e quartis das concentrações dos vários elementos químicos nos musgos observadas em cada campanha.	86

14	Variogramas amostrais das concentrações de Cd nos musgos observadas em cada campanha e respectivos modelos teóricos, estimados pelo método de mínimos quadrados.	89
15	Variogramas amostrais das concentrações de Cr nos musgos observadas em cada campanha e respectivos modelos teóricos, estimados pelo método de mínimos quadrados.	91
16	Variogramas amostrais das concentrações de Cu nos musgos observadas em cada campanha e respectivos modelos teóricos, estimados pelo método de mínimos quadrados.	93
17	Variogramas amostrais das concentrações de Fe nos musgos observadas em cada campanha e respectivos modelos teóricos, estimados pelo método de mínimos quadrados.	96
18	Variogramas amostrais das concentrações de Hg nos musgos observadas em cada campanha. .	97
19	Variogramas amostrais das concentrações de Mn nos musgos observadas em cada campanha e respectivos modelos teóricos, estimados pelo método de mínimos quadrados.	100
20	Variogramas amostrais das concentrações de Ni nos musgos observadas em cada campanha e respectivos modelos teóricos, estimados pelo método de mínimos quadrados.	102
21	Variogramas amostrais das concentrações de Pb nos musgos observadas em cada campanha e respectivos modelos teóricos, estimados pelo método de mínimos quadrados.	104
22	Variogramas amostrais das concentrações de Zn nos musgos observadas em cada campanha e respectivos modelos teóricos, estimados pelo método de mínimos quadrados.	107
23	Variogramas amostrais das concentrações de V nos musgos observadas em cada campanha. . .	108
24	Superfícies preditas das concentrações de As nos musgos observadas em cada campanha, sem partilha de escala.	109
25	Desvios padrão associados a cada valor predito de As, cujas concentrações foram observadas nos musgos em cada campanha.	109
26	Superfícies preditas das concentrações de Cd nos musgos observadas em cada campanha, sem partilha de escala.	110
27	Desvios padrão associados a cada valor predito de Cd, cujas concentrações foram observadas nos musgos em cada campanha.	110
28	Superfícies preditas das concentrações de Cr nos musgos observadas em cada campanha, sem partilha de escala.	111
29	Desvios padrão associados a cada valor predito de Cr, cujas concentrações foram observadas nos musgos em cada campanha.	111
30	Superfícies preditas das concentrações de Cu nos musgos observadas em cada campanha, sem partilha de escala.	112
31	Desvios padrão associados a cada valor predito de Cu, cujas concentrações foram observadas nos musgos em cada campanha.	112
32	Superfícies preditas das concentrações de Fe nos musgos observadas em cada campanha, sem partilha de escala.	113

33	Desvios padrão associados a cada valor predito de Fe, cujas concentrações foram observadas nos musgos em cada campanha.	113
34	Superfícies preditas das concentrações de Mn nos musgos observadas em cada campanha, sem partilha de escala.	114
35	Desvios padrão associados a cada valor predito de Mn, cujas concentrações foram observadas nos musgos em cada campanha.	114
36	Superfícies preditas das concentrações de Ni nos musgos observadas em cada campanha, sem partilha de escala.	115
37	Desvios padrão associados a cada valor predito de Ni, cujas concentrações foram observadas nos musgos em cada campanha.	115
38	Superfícies preditas das concentrações de Pb nos musgos observadas em cada campanha, sem partilha de escala.	116
39	Desvios padrão associados a cada valor predito de Pb, cujas concentrações foram observadas nos musgos em cada campanha.	116
40	Superfícies preditas das concentrações de Zn nos musgos observadas em cada campanha, sem partilha de escala.	117
41	Desvios padrão associados a cada valor predito de Zn, cujas concentrações foram observadas nos musgos em cada campanha.	117
42	Diagramas de extremos e quartis das várias fontes de poluição subjacentes a cada campanha. .	118
43	Variogramas amostrais das fontes de poluição subjacentes à campanha de 1992 e respetivos modelos teóricos, estimados pelo método de mínimos quadrados.	121
44	Variogramas amostrais das fontes de poluição subjacentes à campanha de 1997 e respetivos modelos teóricos, estimados pelo método de mínimos quadrados.	123
45	Variogramas amostrais das fontes de poluição subjacentes à campanha de 2002 e respetivos modelos teóricos, estimados pelo método de mínimos quadrados.	126
46	Variogramas amostrais das fontes de poluição subjacentes à campanha de 2006.	128
47	Desvios padrão associados a cada valor predito das fontes de poluição subjacentes à campanha de 1992.	129
48	Desvios padrão associados a cada valor predito das fontes de poluição subjacentes à campanha de 1997.	129
49	Desvios padrão associados a cada valor predito das fontes de poluição subjacentes à campanha de 2002.	130

Lista de Tabelas

4.1	Parâmetros estatísticos das concentrações dos vários elementos químicos nos musgos observadas em cada campanha.	36
4.2	Estimativas dos parâmetros envolvidos nas estruturas do valor médio e da anisotropia das concentrações de As nos musgos observadas em cada campanha.	38
4.3	Valores dos variogramas empíricos das concentrações de As nos musgos observadas em cada campanha, onde $ N(u) $ representa o cardinal das diferentes classes de distâncias.	38
4.4	Estimativas dos parâmetros envolvidos nas estruturas da associação espacial dos modelos adotados para as concentrações de As nos musgos observadas em cada campanha.	39
4.5	Valores das correlações <i>cofonéticas</i> , para as várias campanhas.	52
4.6	Valores médio dos diferentes grupos subjacentes a cada campanha e respectivos valores de prova do teste de Kruskal-Wallis.	56
4.7	Valores dos determinantes (det) das matrizes de correlação subjacentes a cada campanha e respectivos valores de prova do teste de esfericidade de Bartlett.	57
4.8	<i>Loadings</i> , valores das correlações entre os elementos químicos e as componentes principais e respectivos valores da variância explicada por cada componente principal na campanha de 1992. .	58
4.9	<i>Loadings</i> , valores das correlações entre os elementos químicos e as componentes principais e respectivos valores da variância explicada por cada componente principal na campanha de 1997. .	59
4.10	<i>Loadings</i> , valores das correlações entre os elementos químicos e as componentes principais e respectivos valores da variância explicada por cada componente principal na campanha de 2002. .	60
4.11	<i>Loadings</i> , valores das correlações entre os elementos químicos e as componentes principais e respectivos valores da variância explicada por cada componente principal na campanha de 2006. .	61
5.1	Valores dos indicadores da qualidade do ajustamento dos modelos contendo 1, 2, 3, 4 e 5 fatores das várias campanhas, baseados na estimação dos parâmetros pelo método mínimos quadrados ordinários.	66
5.2	Valores dos <i>loadings</i> das fontes de poluição subjacentes às concentrações dos elementos químicos nos musgos observadas em cada campanha.	68
1	Estimativas dos parâmetros envolvidos nas estruturas do valor médio e da anisotropia das concentrações de Cd nos musgos observadas em cada campanha.	87
2	Valores dos variogramas empíricos das concentrações de Cd nos musgos observadas em cada campanha, onde $ N(u) $ representa o cardinal das diferentes classes de distâncias.	87

3	Estimativas dos parâmetros envolvidos nas estruturas da associação espacial dos modelos adotados para as concentrações de Cd nos musgos observadas em cada campanha.	88
4	Estimativas dos parâmetros envolvidos nas estruturas do valor médio e da anisotropia das concentrações de Cr nos musgos observadas em cada campanha.	89
5	Valores dos variogramas empíricos das concentrações de Cr nos musgos observadas em cada campanha, onde $ N(u) $ representa o cardinal das diferentes classes de distâncias.	89
6	Estimativas dos parâmetros envolvidos nas estruturas da associação espacial dos modelos adotados para as concentrações de Cr nos musgos observadas em cada campanha.	90
7	Estimativas dos parâmetros envolvidos nas estruturas do valor médio e da anisotropia das concentrações de Cu nos musgos observadas em cada campanha.	91
8	Valores dos variogramas empíricos das concentrações de Cu nos musgos observadas em cada campanha, onde $ N(u) $ representa o cardinal das diferentes classes de distâncias.	92
9	Estimativas dos parâmetros envolvidos nas estruturas da associação espacial dos modelos adotados para as concentrações de Cu nos musgos observadas em cada campanha.	93
10	Estimativas dos parâmetros envolvidos nas estruturas do valor médio e da anisotropia das concentrações de Fe nos musgos observadas em cada campanha.	94
11	Valores dos variogramas empíricos das concentrações de Fe nos musgos observadas em cada campanha, onde $ N(u) $ representa o cardinal das diferentes classes de distâncias.	94
12	Estimativas dos parâmetros envolvidos nas estruturas da associação espacial dos modelos adotados para as concentrações de Fe nos musgos observadas em cada campanha.	95
13	Estimativas dos parâmetros envolvidos nas estruturas do valor médio e da anisotropia das concentrações de Hg nos musgos observadas em cada campanha.	96
14	Valores dos variogramas empíricos das concentrações de Hg nos musgos observadas em cada campanha, onde $ N(u) $ representa o cardinal das diferentes classes de distâncias.	97
15	Estimativas dos parâmetros envolvidos nas estruturas do valor médio e da anisotropia das concentrações de Mn nos musgos observadas em cada campanha.	98
16	Valores dos variogramas empíricos das concentrações de Mn nos musgos observadas em cada campanha, onde $ N(u) $ representa o cardinal das diferentes classes de distâncias.	98
17	Estimativas dos parâmetros envolvidos nas estruturas da associação espacial dos modelos adotados para as concentrações de Mn nos musgos observadas em cada campanha.	99
18	Estimativas dos parâmetros envolvidos nas estruturas do valor médio e da anisotropia das concentrações de Ni nos musgos observadas em cada campanha.	100
19	Valores dos variogramas empíricos das concentrações de Ni nos musgos observadas em cada campanha, onde $ N(u) $ representa o cardinal das diferentes classes de distâncias.	100
20	Estimativas dos parâmetros envolvidos nas estruturas da associação espacial dos modelos adotados para as concentrações de Ni nos musgos observadas em cada campanha.	101
21	Estimativas dos parâmetros envolvidos nas estruturas do valor médio e da anisotropia das concentrações de Pb nos musgos observadas em cada campanha.	102

22	Valores dos variogramas empíricos das concentrações de Pb nos musgos observadas em cada campanha, onde $ N(u) $ representa o cardinal das diferentes classes de distâncias.	103
23	Estimativas dos parâmetros envolvidos nas estruturas da associação espacial dos modelos adotados para as concentrações de Pb nos musgos observadas em cada campanha.	104
24	Estimativas dos parâmetros envolvidos nas estruturas do valor médio e da anisotropia das concentrações de Zn nos musgos observadas em cada campanha.	105
25	Valores dos variogramas empíricos das concentrações de Zn nos musgos observadas em cada campanha, onde $ N(u) $ representa o cardinal das diferentes classes de distâncias.	105
26	Estimativas dos parâmetros envolvidos nas estruturas da associação espacial dos modelos adotados para as concentrações de Zn nos musgos observadas em cada campanha.	106
27	Estimativas dos parâmetros envolvidos nas estruturas do valor médio e da anisotropia das concentrações de V nos musgos observadas em cada campanha.	107
28	Valores dos variogramas empíricos das concentrações de V nos musgos observadas em cada campanha, onde $ N(u) $ representa o cardinal das diferentes classes de distâncias.	107
29	Estimativas dos parâmetros envolvidos nas estruturas do valor médio e da anisotropia das fontes de poluição subjacentes à campanha de 1992.	119
30	Valores dos variogramas empíricos das fontes de poluição subjacentes à campanha de 1992, onde $ N(u) $ representa o cardinal das diferentes classes de distâncias.	119
31	Estimativas dos parâmetros envolvidos nas estruturas da associação espacial dos modelos adotados para as fontes de poluição subjacentes à campanha de 1992.	120
32	Estimativas dos parâmetros envolvidos nas estruturas do valor médio e da anisotropia das fontes de poluição subjacentes à campanha de 1997.	121
33	Valores dos variogramas empíricos das fontes de poluição subjacentes à campanha de 1997, onde $ N(u) $ representa o cardinal das diferentes classes de distâncias.	122
34	Estimativas dos parâmetros envolvidos nas estruturas da associação espacial dos modelos adotados para as fontes de poluição subjacentes à campanha de 1997.	122
35	Estimativas dos parâmetros envolvidos nas estruturas do valor médio e da anisotropia das fontes de poluição subjacentes à campanha de 2002.	124
36	Valores dos variogramas empíricos das fontes de poluição subjacentes à campanha de 2002, onde $ N(u) $ representa o cardinal das diferentes classes de distâncias.	124
37	Estimativas dos parâmetros envolvidos nas estruturas da associação espacial dos modelos adotados para as fontes de poluição subjacentes à campanha de 2002.	125
38	Estimativas dos parâmetros envolvidos nas estruturas do valor médio e da anisotropia das fontes de poluição subjacentes à campanha de 2006.	126
39	Valores dos variogramas empíricos das fontes de poluição subjacentes à campanha de 2006, onde $ N(u) $ representa o cardinal das diferentes classes de distâncias.	126

Capítulo 1

Introdução

Enquadramento

O desenvolvimento tecnológico e científico, a que assistimos nos dias de hoje, teve e tem uma grande influência nas concentrações de vários elementos químicos no ambiente. Por esta razão e por se ter verificado os efeitos tóxicos que estes elementos têm para a saúde do ser humano e para outros organismos, torna-se imprescindível a sua avaliação, bem como a monitorização da concentração destes poluentes no meio ambiente. Alguns organismos vivos apresentam capacidades de acumulação de elementos químicos e uma sensibilidade à poluição e, por este motivo, é possível utilizá-los na quantificação das concentrações dos poluentes (Figueira, 2002). Este processo denomina-se por biomonitorização e é um indicador das concentrações dos elementos químicos no ambiente, sendo uma metodologia bastante simples e com uma elevada resolução espacial (correspondendo, no caso em estudo, a Portugal continental) e alguma resolução temporal, que proporciona uma interpretação relevante em termos de efeitos sobre o ambiente.

De forma a compreender a importância de cada fonte de poluição ambiental, devem-se analisar as relações entre as concentrações dos elementos químicos nos musgos e os diversos parâmetros ambientais, geológicos e antropogénicos (Menezes, 2007). Verifica-se, assim, uma necessidade contínua de obter modelos estatísticos adequados aos dados em estudo com um rigor cada vez mais elevado (Soares, 2000). O objetivo desta análise é prever a concentração dos elementos no meio ambiente para os diversos pontos geográficos de uma dada região.

Objetivos

Este trabalho é resultado da análise estatística das concentrações de vários elementos químicos nos musgos, tais como As, Cd, Cr, Cu, Fe, Hg, Mn, Ni, Pb, Zn e V. Os musgos foram recolhidos em quatro campanhas, que ocorreram ao longo de vários meses, pelo Centro de Biologia Ambiental da Universidade de Lisboa, através de uma rede de biomonitorização abrangendo Portugal continental. Nas campanhas de 1992, 1997 e 2002 foram recolhidas amostras de musgo em 173, 177 e

151 localizações, respetivamente. Na campanha de 2006, com características distintas das anteriores, consideraram-se apenas 98 localizações contidas na região Centro de Portugal (NUT II).

Com os resultados da aplicação do sistema de biomonitorização ambiental é possível o emprego de metodologias estatísticas, nomeadamente multivariadas, permitindo a identificação das interações do organismo com os parâmetros ambientais. Neste sentido, torna-se indispensável a compreensão das principais características e comportamentos de cada um dos elementos químicos, a reconhecimento dos locais mais afetados pelas elevadas concentrações dos vários elementos e o reconhecimento das relações lineares existentes entre os conjuntos múltiplos de dados.

Uma vez que as concentrações desses elementos químicos podem resultar de várias fontes de contaminação ambiental, como por exemplo antropogénicas, geogénicas e geológicas (Figueira et al., 2002), é fundamental estabelecer relações entre os vários elementos químicos e essas fontes de poluição. Desta forma, um dos objetivos deste trabalho é a identificação das diferentes fontes de poluição subjacentes às concentrações dos elementos químicos nos musgos, constituindo variáveis que não podem ser medidas diretamente no meio ambiente, designadas por variáveis latentes. Consequentemente, recorre-se ao método estatístico de análise fatorial, o qual permite que estas variáveis possam ser tratadas como contínuas, resultando da separação de um número limitado de associações discretas, cujo método fornece um meio para testar se esta divisão é estatisticamente significativa.

Um outro objetivo deste estudo é a identificação e a validação de modelos, que possibilitem a reconstrução da distribuição espacial de cada fonte de poluição ao longo de uma determinada região geográfica. Nesta perspetiva, com o auxílio da geoestatística é possível fazer a estimação dos valores das concentrações das variáveis latentes para locais não amostrados, com base nos valores reais observados, permitindo a construção de mapas que, posteriormente, podem ser utilizados para a avaliação da qualidade ambiental. Com a aplicação desta metodologia, denominada por *kriging*, que não é mais do que um método de interpolação de um modelo probabilista de inferência espacial, tem-se acesso ao erro de interpolação envolvido, podendo ser utilizado como uma medida de incerteza das predições.

Este tipo de raciocínio permitir-nos-á avaliar a qualidade ambiental através de um resumo da informação disponível das concentrações dos vários elementos químicos nos musgos, salientando-se simultaneamente os locais que padecem de determinada fonte de poluição.

Organização

A metodologia descrita na secção anterior teve como principal objetivo a integração coerente dos diversos métodos estatísticos, utilizados nas diferentes fases deste trabalho. Assim, para além deste capítulo de introdução, os capítulos 2 e 3 referem-se, respetivamente, à apresentação dos dados analisados e das metodologias estatísticas usadas. Os capítulos 4 e 5 são relativos à apresentação dos resultados da análise estatística dos dados, incluindo a discussão e a interpretação dos mesmos.

O segundo capítulo dá ênfase ao fenómeno da poluição, sendo apresentado um pequeno resumo do que é um sistema de biomonitorização ambiental, metodologia adotada para a recolha dos dados disponíveis neste estudo. Abordam-se, ainda, algumas características dos elementos químicos amostrados ao longo das quatro campanhas.

No terceiro capítulo apresentam-se as metodologias estatísticas utilizadas no âmbito deste trabalho, fazendo-se uma pequena exposição das bases teóricas, quer da análise estatística multivariada, quer da análise estatística espacial, mais especificamente da geoestatística.

No quarto capítulo apresentam-se os resultados obtidos da análise individual dos elementos químicos, contendo uma análise descritiva univariada e uma análise da variabilidade espacial de cada um dos elementos. Contém, ainda, a análise multivariada dos dados, identificando-se os locais mais afetados pelas elevadas concentrações dos elementos e as relações subjacentes aos diversos elementos químicos.

O quinto capítulo corresponde à identificação das principais fontes de poluição através da análise fatorial, cujo objetivo final é a obtenção de mapas de valores preditos das diferentes fontes de poluição.

Num último capítulo, apresentam-se as principais conclusões decorrentes deste estudo com a aplicação das análises estatísticas, indicando alguns comentários acerca dos resultados obtidos e dos trabalhos que podem ser realizados posteriormente.

Capítulo 2

Componentes do Sistema Ambiental

A poluição na Europa tem vindo a aumentar drasticamente desde a Revolução Industrial. Muitos poluentes eram, e são ainda hoje, libertados para a atmosfera através de diversos meios (por exemplo com a queima do carvão mineral, com o tráfego automóvel, com atividades industriais, etc.) (Anderberg et al., 2000). A saúde do Homem é muito afetada com este tipo de poluição, nomeadamente com a presença de alguns elementos químicos (arsénio, cádmio, crómio, chumbo, manganês, mercúrio, etc.) encontrados com regularidade no ar ambiente (McGranahan and Murray, 2003). Este facto é evidenciado através da ocorrência de doenças respiratórias como bronquite, rinite alérgica, alergias, asma, irritação da pele, lacrimação exagerada, etc. No entanto, não é só o ser humano a sofrer com a poluição, esta também tem influenciado os ecossistemas e o património histórico e cultural através do efeito de estufa e da chuva ácida.

A consciencialização dos problemas associados à poluição, nomeadamente com a contaminação da água e do solo, e a preocupação com os efeitos tóxicos dos poluentes nos organismos vivos fez com que, na última metade do século XX, fosse necessária a quantificação dos poluentes presentes no sistema ambiental (Hill, 2010), que tem vindo a ser analisada quer a nível governativo, quer a nível científico. Desta forma, tornou-se indispensável o estabelecimento de uma rede de amostragem em Portugal continental, existindo no entanto, um número muito reduzido de estações fixas equipadas para a quantificação da deposição atmosférica dos elementos químicos (Figueira, 2002). Devido aos custos elevados daí resultantes, houve a necessidade de usar organismos vivos para a monitorização ambiental - biomonitorização, que pode ser um complemento à monitorização física.

A informação relevante da biomonitorização é geralmente determinada através das alterações do comportamento dos biomonitores ou das concentrações de determinadas substâncias nos tecidos dos mesmos. Uma das vantagens que a biomonitorização apresenta relativamente à monitorização física diz respeito aos custos, uma vez que os instrumentos de monitorização são onerosos. No entanto, é fundamental a noção da existência de desvantagens inerentes à utilização de biomonitores em vez de sistemas físicos. Por exemplo, a reação dos biomonitores não depende apenas da quantidade dos elementos no meio envolvente, mas também dos parâmetros do biomonitor (idade, espécie, estado de vitalidade, etc.) e de parâmetros ambientais (tipo de solo, precipitação, humidade, topografia,

temperatura, etc.), sendo que as leituras obtidas nos biomonitores apresentam maior variabilidade do que as dos instrumentos, cuja consequência é a obtenção de resultados com uma interpretação mais complexa. Outras desvantagens são a absorção incompleta dos elementos químicos que apresentam menor afinidade com o organismo vivo e a dificuldade da escolha de condições ótimas na amostragem (Menezes, 2007).

Porém, de acordo com Figueira (2002), existe uma maior relevância das vantagens versus as desvantagens. Assim, este método começou a ser reconhecido e recomendado por entidades nacionais e internacionais de alguns países, como Alemanha, Inglaterra, Noruega, Polónia, Espanha, Itália, entre outros (Whitfield, 2001; Berg and Steinnes, 1997; Grodzinska et al., 1999; Fernández et al., 2000; Gerdol et al., 2000).

Um estudo de biomonitorização procura reconhecer as tendências existentes na resposta dos organismos às alterações do ambiente, ao longo do espaço e do tempo. Neste sentido, existe uma abordagem muito própria num trabalho de biomonitorização - sistema de biomonitorização ambiental, correspondendo a um conjunto de passos a desenvolver ao longo do trabalho. Apresentam-se, na Figura 2.1, as relações existentes entre cada uma dessas etapas.

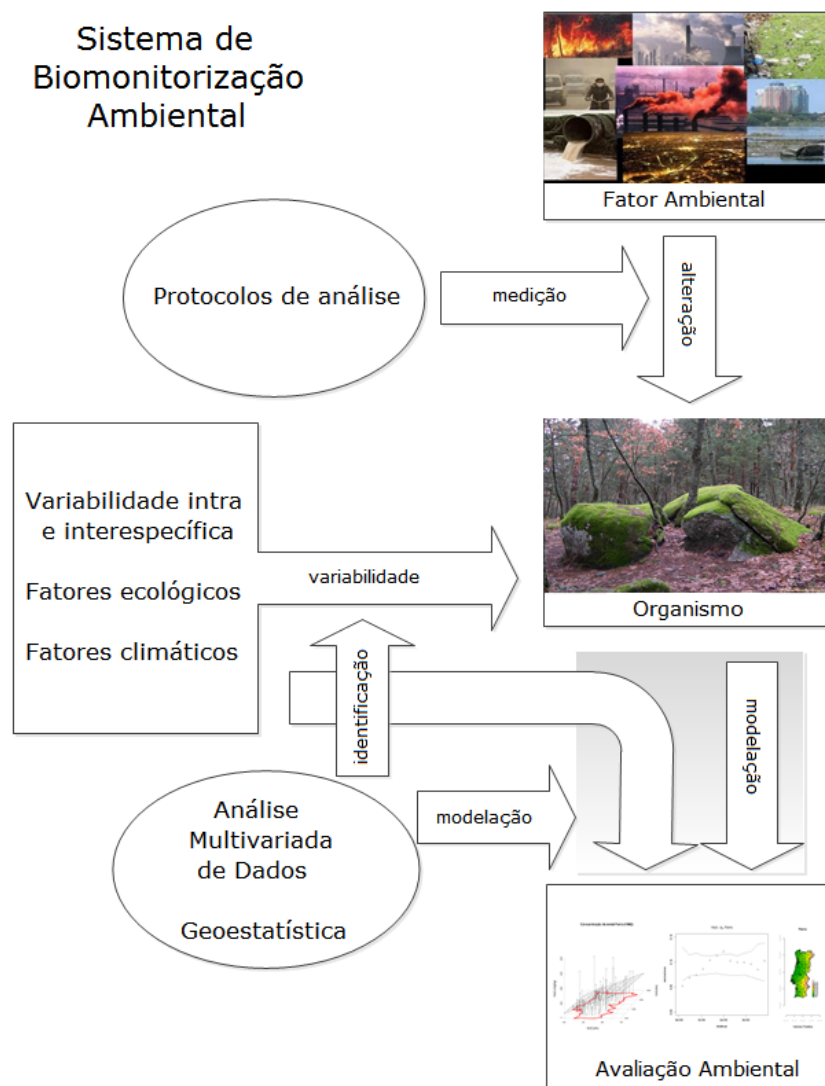


Figura 2.1: Esquema de um estudo de biomonitorização, adaptado de Figueira (2002).

A aplicação deste sistema implica a realização de um estudo aprofundado das características morfológicas e fisiológicas, bem como a realização de uma avaliação dos parâmetros ambientais que intervêm na resposta do organismo ao ambiente. Estes estudos devem permitir o desenvolvimento de metodologias de análise estatística que possibilitem a identificação das interações do organismo com o ambiente, bem como a sua descrição e modelação espacial e temporal. A utilização de métodos da análise multivariada possibilita a identificação das relações lineares entre os conjuntos de dados multivariados, de forma a encontrar os grupos de parâmetros do organismo e do ambiente que se encontram relacionados. O uso da geoestatística permite a determinação e modelação da continuidade, no espaço e/ou no tempo, dos processos associados à resposta do organismo ao ambiente, obtendo-se mapas de valores preditos passíveis de serem utilizados para a avaliação da qualidade ambiental das áreas estudadas (Figueira, 2002).

2.1 Estado da arte

Apesar das concentrações dos elementos químicos variarem com determinados fatores, tais como o tipo de superfície, o tempo de exposição, a tendência de absorverem determinado elemento por processos biológicos, a capacidade de troca iónica, entre outros, a quantidade desses elementos em organismos vivos está estritamente relacionada com as concentrações de partículas existentes na atmosfera envolvente (Sérgio et al., 1993). Em Portugal, esses organismos vivos têm sido utilizados como sensores da qualidade do ar e solo, cujos estudos estão referenciados nesta secção.

Relativamente ao sistema de biomonitorização ambiental

Um dos primeiros trabalhos nesta área em Portugal foi realizado por Sérgio et al. (1993). Este trabalho tinha como objetivo a quantificação de elementos químicos em Portugal, por meio da análise de briófitos. Estes autores, após tratamento estatístico e cartográfico, concluíram que existia uma grande dispersão dos valores e que, na generalidade, ao comparar-se a distribuição regional dos diferentes elementos foram encontradas analogias entre o chumbo e o zinco, em que a ação antropogénica é a causa mais evidente para esta relação; entre o cádmio e o crómio, em que a presença de minas e agentes erosivos parecem ter mais peso nesta associação; e entre o ferro e o manganês, em que a componente geológica deve ser a fonte de contaminação mais importante.

Figueira (2002) termina o seu doutoramento cujo objetivo era definir uma metodologia de análise ambiental por biomonitorização, baseando-se na identificação e caracterização da variabilidade inerente ao biomonitor e das relações entre o biomonitor e o ambiente. Neste estudo, calcularam-se mapas de interpolação com base na análise da estrutura espacial das variáveis. Assim, para cada metal foi ajustado um modelo para o variograma e aplicado *kriging* ordinário para proceder à predição dos valores em localizações não amostradas. A partir de informações de campo, cartografia e outras informações disponíveis foi possível a classificação de cada um dos locais de amostragem em relação

a várias características ambientais, antropogénicas, litológicas e pedológicas. Os mapas produzidos permitiram a identificação das áreas de maior deposição dos elementos, bem como a distribuição média destes. Segundo Figueira (2002), os elementos analisados através dos musgos apresentam associações que resultam de fontes de contaminação comuns, porém estas podem expressar-se de forma diferente quando se toma em atenção a variabilidade espacial.

Após a realização destes dois trabalhos pioneiros em Portugal, surge uma nova base de dados analisada por Figueira et al. (2004), tendo como objetivo a apresentação dos dados, bem como a discussão de padrões das distribuições dos elementos químicos e eventuais fontes de emissão e contaminação. Foram identificadas associações entre os vários elementos e as relações entre estes e os diversos tipos de contaminação ambiental através da análise em componentes principais. Posteriormente, foram apresentados mapas de distribuição dos elementos calculados a partir da estimação geoestatística, após caracterização dos variogramas amostrais.

Posteriormente, surge a dissertação de Menezes (2007) para obtenção do grau de mestre. O principal objetivo deste estudo foi avaliar a capacidade de diferentes metodologias geoestatísticas, nomeadamente na escolha de diferentes modelos para a estrutura de correlação espacial, com o intuito de resolver problemas associados à qualidade dos dados de amostragem sendo, para tal, utilizado *kriging* da indicatriz. Posteriormente, fizeram a modelação geoestatística das variáveis através da análise da continuidade espacial e do *kriging*. Assim, foram determinados os erros de estimação relativos às diferentes metodologias utilizadas, aferindo-se quais as mais adequadas para os diferentes casos.

Um novo trabalho realizado por Reis et al. (2007) teve como objetivo a identificação de localizações que apresentam valores elevados de metais oriundos de resíduos perigosos, devido à extração mineira em minas portuguesas que já não operam. A mina de Lousal encontrava-se encerrada, mas os resíduos dos metais pesados permaneciam no local. Aplicando a geoestatística, com o intuito de diagnosticar a área potencialmente contaminada, eles pretenderam avaliar a qualidade do solo através da sua vulnerabilidade. Para proceder a esta avaliação, construíram mapas probabilísticos do elemento químico arsénio, mostrando as probabilidades de contaminação associadas à área em questão. Estes mapas foram, posteriormente, transformados em mapas de incerteza.

Um outro estudo realizado, também, por Reis et al. (2010) pretendeu identificar os padrões espaciais de alguns metais, em solos e musgos, na região central de Portugal. Os objetivos deste foram identificar as relações entre cinco elementos químicos (cobre, chumbo, zinco, crómio e arsénio) em três meios diferentes (solo, solo de fundo e briófitos), tendo em atenção algumas características específicas do local, através da análise de correspondência múltipla; definir padrões espaciais para as associações identificadas por análise de correspondência múltipla usando a variografia e o *kriging*; e, ainda, avaliar a deposição atmosférica dos elementos no solo, cruzando os resultados com os dos biomonitores. Os resultados indicaram concentrações relativamente baixas dos elementos químicos, quer em solos quer em musgos.

Martins (2011) concluiu a sua dissertação para obtenção do grau de mestre em Engenharia do

Ambiente, cujo objetivo consistiu na produção de um indicador espacial da variação temporal da deposição atmosférica de cobre determinada em biomonitores, com base na interpolação espaço-temporal por *kriging* da indicatriz, propondo, desta forma, uma metodologia para a avaliação da evolução da qualidade do ar baseada nas funções de distribuição locais estimadas. Este estudo demonstrou o como o *kriging* da indicatriz pode ser útil no cálculo das funções das distribuições de probabilidade locais de um conjunto de dados, sendo que, através do método de interpolação espaço-temporal, conseguiram representar a evolução espaço-temporal da contaminação atmosférica de cobre em Portugal dos últimos 20 anos.

Relativamente às metodologias estatísticas aplicadas

Os principais objetivos científicos da geoestatística são a estimação e a predição. Para tal é bastante útil o cálculo do variograma empírico, tratando-se de uma ferramenta crucial na investigação espacial e no processo da modelação da covariância. Porém, por vezes torna-se difícil identificar as suas propriedades, num contexto de exploração das características dos conjuntos de dados individuais. Bowman and Crujeiras (2013) sugerem que a adequação das hipóteses, como a isotropia e a estacionaridade, podem ser investigadas através da construção de estatísticas de teste apropriadas e de distribuições exigidas para o cálculo dos valores de prova associados, recorrendo a métodos quadráticos.

Os modelos espaço-temporais são modelos que surgiram recentemente e que constituem uma área com grande potencial para a pesquisa científica. Existem várias referências a estes métodos, mas muitos destes não estão implementados computacionalmente. Ribeiro and Mingoti (2006) propõem um programa desenvolvido em R, com o auxílio da *package* *geoR*, que implementa uma versão da metodologia proposta por Kyriakidis and Journel (1999), mostrando que este pode ser utilizado para fazer previsões para os tempos nos quais o processo estocástico subjacente aos dados não fora amostrado.

Militino et al. (2008) apresenta um novo estudo no qual é proposto um modelo linear misto, que incorpora tanto uma componente espacial, bem como informações longitudinais, com o objetivo específico de deteção de áreas com excesso de nitrato nas águas subterrâneas. O modelo proposto neste trabalho, estende-se de uma forma natural para a abordagem tradicional do *kriging*, incorporando a dimensão longitudinal.

2.2 Elementos químicos analisados

Neste trabalho foram analisados onze elementos químicos: arsénio (As), cádmio (Cd), crómio (Cr), cobre (Cu), ferro (Fe), mercúrio (Hg), manganês (Mn), níquel (Ni), chumbo (Pb), zinco (Zn) e vanádio (V), cujos mapas dos valores das concentrações dos vários elementos químicos nos musgos estão contidos no Anexo II.

De seguida apresentam-se os elementos químicos referidos, referindo as suas principais características quer geogénicas, quer antropogénicas.

Arsénio (As) O As é um elemento que, no meio ambiente, existe normalmente combinado com outros elementos apresentando-se na forma de compostos orgânicos e inorgânicos, oriundos de fontes geogénicas e/ou antropogénicas. Por exemplo, o As pode surgir na água potável, tendo origem nas nascentes, e em poeiras. Mas o facto é que também aparece em resultado da poluição industrial, através dos depósitos finais de rejeitos químicos, da manufatura (fusão) de cobre, da indústria mineira, da queima de combustíveis fósseis (nomeadamente, da combustão do carvão) (Faus-Kessler et al., 2000) e da indústria química de pesticidas e herbicidas (Rhoades, 1999). O As, quando encontrado na sua forma pura, é um elemento químico altamente tóxico (Ferreira, 2004). No entanto, até uma determinada concentração, este elemento não é prejudicial à saúde do ser humano.

Cádmio (Cd) O Cd é um elemento químico presente na crosta terrestre, cujas fontes geogénicas na atmosfera surgem com a atividade vulcânica, a erosão de rochas sedimentares e fosfáticas e os incêndios florestais (Fergusson, 1990). Quando este elemento atinge concentrações elevadas, torna-se num elemento tóxico (Sérgio et al., 1993). As fontes antropogénicas deste elemento são as atividades de mineração, produção, consumo e disposição de produtos que utilizam Cd (como baterias de níquel-cádmio, estabilizadores de produtos de PVC (*policloreto de vinil*), ligas de cádmio e componentes eletrónicas, fertilizantes, etc.).

Crómio (Cr) O Cr é considerado um elemento químico com um índice de bioacumulação regrado, sendo moderadamente tóxico quer para os animais quer para as plantas (Ferreira, 2004). A existência deste elemento encontra-se tipicamente associado ao transporte erosivo do solo por ação do vento (Menezes, 2007). Este elemento pode ter, ainda, origem nas indústrias metalúrgicas e refinarias, bem como nas indústrias de vidro, cimenteiras e extrações mineiras (Faus-Kessler et al., 2000). A concentração de Cr é menor em rochas graníticas, aumentando o seu valor nas rochas ultramáficas/básicas.

Cobre (Cu) O Cu é encontrado na natureza na forma de calcopirita CuFeS_2 , sendo um micronutriente necessário a todos os organismos vivos (Sérgio et al., 1993). No entanto, quando encontrado com concentrações elevadas, este é considerado um metal pesado, apresentando um índice de bioacumulação elevado (Ferreira, 2004). O Cu, em contacto com ar atmosférico por tempo prolongado, sofre oxidação formando numa superfície uma película tóxica originada a partir de uma mistura de óxidos, hidróxidos e carbonatos. A origem deste elemento é essencialmente antropogénica (Szczepaniak and Biziuk, 2003), nomeadamente nas indústrias de processamento de metais e no tráfego rodoviário (Faus-Kessler et al., 2000).

Ferro (Fe) O Fe apresenta um baixo índice de bioacumulação (Ferreira, 2004). Este é um dos metais mais comuns na natureza, sendo um micronutriente muito pouco tóxico (Sérgio et al., 1993). Este ele-

mento encontra-se fortemente associado ao transporte erosivo do solo por ação do vento, bem como à atividade industrial de processamento de ferro e aço (Menezes, 2007). O Fe constitui, juntamente com o carvão e o petróleo, a matéria prima mais importante da economia mundial. Apesar de atualmente este estar a ser substituído em grande parte pelos metais leves e plásticos, a procura deste metal, assim como a sua produção a nível mundial, aumentam constantemente.

Mercúrio (Hg) O Hg é um metal pesado em estado líquido absorvido pelos organismos vivos, presente no meio ambiente sob a forma de compostos orgânicos e inorgânicos (Harrison, 2001). Por contaminação da água ou do solo, insere-se com facilidade na cadeia alimentar, sendo por isso, responsável por vários casos de intoxicação humana e animal. Este metal é largamente usado nos diversos ramos da atividade humana como mineração (amalgamento do ouro), indústrias cerâmicas e farmacêuticas, instrumentos de medição, etc. (Harrison, 2001). Está também referenciado pela gaseificação da crosta terrestre, erupções vulcânicas e pela queima de combustíveis fósseis.

Manganês (Mn) O Mn é um metal pesado que apresenta um baixo índice de bioacumulação (Ferreira, 2004). Em comparação com outros metais, é um dos elementos metálicos menos tóxico (Sérgio et al., 1993), podendo encontrar-se associado à proximidade da costa (Rhoades, 1999). As principais fontes deste metal são geogénicas, por exemplo através da meteorização de rochas. No entanto, este também pode estar associado à fabricação de aços e ligas metálicas (como o alumínio). Um outro composto de Mn, *Metilciclopentadienil Manganês Tricarbonil*, é utilizado como aditivo na gasolina sem chumbo para aumentar o índice de octanas e, conseqüentemente, reduzir os efeitos de explosão do motor de um automóvel.

Níquel (Ni) O Ni é um metal pesado que apresenta um nível de bioacumulação moderado (Ferreira, 2004), dependendo a sua toxicidade dos compostos formados (Menezes, 2007). Este metal tem origem geogénica através do transporte erosivo do solo por ação do vento (Szczepaniak and Biziuk, 2003). No entanto, este também apresenta fontes antropogénicas através de atividades industriais sobretudo metalúrgicas, centrais térmicas e combustão do carvão. Este metal é um elemento químico com grande volatilidade (Sérgio et al., 1993), o que conduz à sua disseminação a uma longa distância.

Chumbo (Pb) O Pb é um metal pesado, não essencial ao ser humano, que apresenta um índice de bioacumulação moderado (Ferreira, 2004). É considerado um elemento tóxico para a maioria dos organismos vivos, sobretudo no que se refere às suas formas iónicas solúveis (Menezes, 2007), proveniente de fontes antropogénicas (Sérgio et al., 1993). O mais amplo uso deste elemento químico é na fabricação de acumuladores, sendo também utilizado na fabricação de forros para cabos, na construção civil, etc. (Ferreira, 2004). A fabricação de Pb *tetra etílico* tem diminuído muito devido a regulamentações ambientais cada vez mais restritas no mundo, no que diz respeito à sua principal aplicação de aditivo na gasolina (Faus-Kessler et al., 2000).

Zinco (Zn) O Zn é considerado um metal pesado de baixa toxicidade com um índice de bioacumulação razoavelmente alto (Ferreira, 2004). Quando este metal se encontra acima de determinados valores, causa alterações no crescimento e produtividade do musgo *Hypnum cupressiforme* (Sérgio et al., 1993). As origens deste metal podem ser antropogénicas devido à indústria transformadora, metalúrgica e às extrações mineiras de Zn (Faus-Kessler et al., 2000), bem como ao tráfego automóvel (Rhoades, 1999); ou ainda geogénicas como as poeiras. Este metal apresenta-se em muitos casos associado ao Cd, e como este é igualmente volátil, a sua dispersão pode ir a uma longa distância (Sérgio et al., 1993).

Vanádio (V) O V é considerado um elemento químico tóxico, apresentando um índice de bioacumulação muito baixo (Ferreira, 2004). Alguns dos compostos deste metal pesado são utilizados como catalisadores na indústria química (Faus-Kessler et al., 2000). As origens deste metal são essencialmente geogénicas, encontrando-se tipicamente associado ao transporte erosivo do solo por ação do vento (Szczepaniak and Biziuk, 2003). A indústria do petróleo e refinarias bem como a produção de aço são origens antropogénicas possíveis deste elemento (Ferreira, 2004).

2.3 Amostragem

Os dados relativos às concentrações dos elementos químicos nos musgos foram observados em todo o território de Portugal continental, cujos distritos e regiões da NUT II estão apresentados no Anexo I.

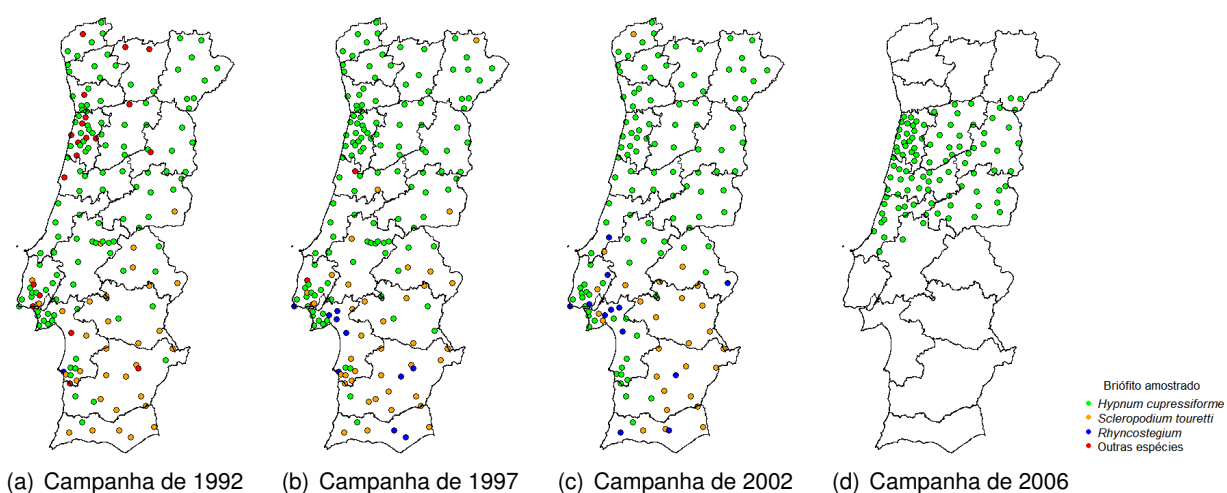


Figura 2.2: Mapas de Portugal continental com as diferentes localizações amostradas em cada campanha.

Nas campanhas de 1992, 1997 e 2002 foram recolhidas amostras de musgo em 173, 177 e 151 localizações, respetivamente (Figura 2.2). Os briófitos selecionados como elementos de biomonitorização foram o *Hypnum cupressiforme*, sendo um musgo cosmopolita e ubiquista frequente na região a norte do rio Tejo (Figueira et al., 2004), e *Scleropodium touretti*, que é uma espécie submediterrânea resistente a maior secura e insolação, sendo mais frequente a sul do país (Figueira et al., 2002).

Na campanha de 1992 também foram recolhidos os briófitos *Rhynchostegium* e *Frullania dilatata*. Por último, na campanha de 2006, com características distintas das anteriores, consideraram-se apenas 98 localizações contidas na região Centro de Portugal, sendo apenas amostrado o musgo *Hypnum cupressiforme*. O desenho para estes dados é não uniforme uma vez que a fração da amostragem varia sistematicamente ao longo da região em estudo.

As amostras da campanha de 1992 e de 1997 foram recolhidas entre Janeiro de 1991 e Março de 1992 e entre Novembro de 1995 e Março de 1999, respetivamente, abrangendo a totalidade do país numa rede de $30 \times 30 \text{ km}$ (Sérgio et al., 1993). Esta rede foi reduzida para uma grelha de $10 \times 10 \text{ km}$ nas áreas de maior densidade populacional ou implantação industrial, aplicada às áreas urbanas de Lisboa e Vale do Tejo, Porto-Aveiro e Complexo Industrial de Sines, seguindo as diretivas da União Europeia (Figueira, 2002). As amostras de musgo da terceira campanha foram recolhidas entre Março de 2001 e Outubro de 2002, onde alguns dos locais amostrados são coincidentes com as campanhas anteriormente efetuadas em 1992 e 1997. Na quarta campanha, as amostras de musgo foram distribuídas por uma rede de amostragem de $5 \times 5 \text{ km}$. Esta rede encontra-se dentro das recomendações propostas por Fernández et al. (2005) para estudos europeus de biomonitorização, decorrendo entre Março e Junho de 2006, existindo mais uma amostra decorrente de Abril de 2004.

Os musgos foram recolhidos com o utensílio de luvas e sacos de plástico para evitar a contaminação metálica do material e também pelo manuseamento. Os locais amostrados, sempre que possível, foram afastados mais de 500 m das estradas principais e áreas habitáveis e mais de 100 m de habitações rurais. Os detalhes sobre o processo de amostragem e as técnicas de medição utilizadas são descritas detalhadamente por Figueira et al. (2002) e Martins (2011).

Capítulo 3

Enquadramento Teórico

Neste capítulo apresenta-se o enquadramento teórico das metodologias utilizadas no que diz respeito a técnicas de análise estatística multivariada e análise geoestatística.

Neste sentido, faz-se uma abordagem às metodologias: análise de *clusters*, análise em componentes principais e análise fatorial, que têm por objetivo a análise e a compreensão de matrizes de dados p -dimensionais, permitindo a extração das variações dominantes e reveladoras de tendências comuns, através da análise da variabilidade existente num conjunto de dados.

Apresentam-se também as metodologias relativas à análise geostatística. Este ramo ocupa, nos dias de hoje, um lugar importante devido ao desenvolvimento tecnológico que permite a fácil obtenção de dados geo-referenciados. Os modelos de análise geoestatística têm por base dados recolhidos em localizações espaciais distintas, medindo a relação entre as observações amostradas em diversos locais. Esta relação deve representar a noção intuitiva da existência de uma correlação entre dados situados numa proximidade espacial, permitindo observar a forma de como variáveis obtidas em pontos distintos do espaço se relacionam.

3.1 Métodos de análise multivariada

Neste tipo de abordagem, dispõe-se de uma matriz de dados, com dimensão $n \times p$,

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1j} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2j} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ x_{i1} & x_{i2} & \cdots & x_{ij} & \cdots & x_{ip} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nj} & \cdots & x_{np} \end{pmatrix} \quad (3.1)$$

onde n representa o número total de indivíduos, p representa o número de variáveis observadas em cada indivíduo e x_{ij} denota o valor da j -ésima variável aleatória observada no indivíduo i , $j = 1, \dots, p$

e $i = 1, \dots, n$. A matriz \mathbf{X} pode ser considerada como n realizações de um vetor aleatório p -variado \mathbf{x} , onde $\mathbf{x}^T = [X_1, X_2, \dots, X_p]$.

Dado um conjunto de p variáveis aleatórias, X_1, X_2, \dots, X_p , tem-se que

- o vetor valor médio é dado por $\boldsymbol{\mu}^T = [\mu_1, \mu_2, \dots, \mu_p]$. Uma estimativa para $\boldsymbol{\mu}$, baseada no estimador usual de máxima verosimilhança de cada parâmetro $\mu_j = \mathbb{E}[X_j]$, \bar{X}_j , é dada por $\bar{\mathbf{x}}^T = [\bar{x}_1, \bar{x}_2, \dots, \bar{x}_p]$, onde \bar{x}_j é o valor observado da média amostral da variável X_j ;
- o vetor da variância populacional é definido por $\boldsymbol{\sigma}^{2T} = [\sigma_1^2, \sigma_2^2, \dots, \sigma_p^2]$, onde $\sigma_j^2 = \text{Var}[X_j] = \mathbb{E}[(X_j - \mu_j)^2]$. Uma estimativa para $\boldsymbol{\sigma}^2$, baseada no estimador usual de $\text{Var}[X_j]$, $s_j^2 = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2$, é dada por $\mathbf{s}^{2T} = [s_1^2, s_2^2, \dots, s_p^2]$, onde s_j^2 é a variância amostral de X_j ;
- a matriz de covariância populacional é representada por $\boldsymbol{\Sigma}$, onde cada elemento $\sigma_{jk} = \text{Cov}[X_j, X_k] = \mathbb{E}[(X_j - \mu_j)(X_k - \mu_k)]$ é estimado por $s_{jk} = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)$, $j, k = 1, \dots, p$. A matriz formada pelos estimadores s_{jk} é designada por matriz de covariância amostral e é representada por \mathbf{S} .

A matriz de covariância possui propriedades muito importantes e fundamentais para a análise estatística multivariada, nomeadamente o ser simétrica e definida positiva (caso não exista colinearidade entre as variáveis que a definem). No caso da matriz $\boldsymbol{\Sigma}$ apresentar ordem $p \times p$, a propriedade anterior garante a existência de p valores próprios reais positivos distintos associados a vetores próprios ortogonais (Jobson, 1992).

Na sequência da standardização dos dados, obtém-se a matriz de correlação \mathbf{P} , onde o termo da matriz é o coeficiente de correlação $\rho_{jk} = \text{Corr}[X_j, X_k] = \frac{\sigma_{jk}}{\sqrt{\sigma_{jj}\sigma_{kk}}}$, $j, k = 1, \dots, p$. Uma estimativa para \mathbf{P} é a matriz de correlação empírica \mathbf{R} , onde cada elemento $r_{jk} = \frac{1}{n-1} \sum_{i=1}^n \frac{x_{ij} - \bar{x}_j}{s_j} \frac{x_{ik} - \bar{x}_k}{s_k}$ corresponde ao coeficiente de correlação de *Pearson*. À semelhança da matriz de covariância, também esta matriz é simétrica e definida positiva.

3.1.1 Análise de *clusters*

O principal objetivo da análise de *clusters* (AC) é o agrupamento dos indivíduos em função da informação existente, de tal forma que indivíduos pertencentes a um mesmo grupo sejam tão semelhantes quanto possível e sempre mais semelhantes aos elementos do mesmo grupo do que a elementos dos restantes grupos (Chatfield and Collins, 1980). Existem situações em que tanto interessa agrupar os indivíduos como as variáveis, sendo estes dois procedimentos duais no sentido em que os métodos abordados nesta secção tanto operam sobre um conjunto de dados iniciais \mathbf{X} como sobre \mathbf{X}^T . Nesta subsecção refere-se apenas o método hierárquico (Everitt and Hothorn, 2011), existindo outros como por exemplo o método *k-means*, descritos detalhadamente por Gordon (1999, 1987) e Everitt et al. (2011). De acordo com Johnson and Wichern (2007), a AC tem como intuito descobrir agrupamentos

"naturais" de indivíduos. Neste sentido, torna-se imprescindível a definição de uma escala quantitativa para medir as (dis)semelhanças entre eles e a caracterização de um algoritmo de classificação.

O primeiro passo da AC é o cálculo de uma matriz de dissemelhanças inter-individual. No caso de variáveis quantitativas, a medida de dissemelhança mais conhecida é a distância euclidiana (Everitt and Hothorn, 2011), cuja dissemelhança entre os indivíduos i e m é dada por

$$d_{im} = [(\mathbf{x}_i - \mathbf{x}_m)^T (\mathbf{x}_i - \mathbf{x}_m)]^{\frac{1}{2}}, \quad (3.2)$$

onde $i, m = 1, \dots, n$ e \mathbf{x}_i e \mathbf{x}_m são dois vetores de observações da matriz \mathbf{X} relativos aos indivíduos i e m . Por vezes, esta dissemelhança não é satisfatória, sobretudo se as variáveis são observadas em unidades de medida diferentes, se têm variâncias muito distintas ou, ainda, se são correlacionadas. Uma solução para este problema consiste em definir uma distância euclidiana ponderada entre os objetos i e m (Timm, 2002), introduzindo uma matriz de pesos \mathbf{A} da seguinte forma

$$d_{im} = [(\mathbf{x}_i - \mathbf{x}_m)^T \mathbf{A} (\mathbf{x}_i - \mathbf{x}_m)]^{\frac{1}{2}}. \quad (3.3)$$

Segundo Jobson (1992), pode-se, também, construir dissemelhanças com base na família de métricas de *Minkowski*

$$d_{im} = \left[\sum_{j=1}^p |x_{ij} - x_{mj}|^\lambda \right]^{\frac{1}{\lambda}}, \quad (3.4)$$

onde $\lambda \geq 1$. No caso de $\lambda = 1$ tem-se a medida de dissemelhança de *city-block*; se $\lambda = 2$ tem-se a medida de dissemelhança euclidiana e, ainda, se $\lambda \rightarrow \infty$ obtém-se a medida de dissemelhança máxima ou de *Chebychev*, dada por $d_{im} = \max_j |x_{ij} - x_{mj}|$.

Depois de definida a medida de dissemelhança, segue-se a escolha do método de agrupamento. Utilizando métodos hierárquicos, os grupos formam uma hierarquia caracterizada pelo facto de dado dois quaisquer grupos, estes ou são disjuntos ou um deles está contido no outro. Para aplicar estes métodos recorre-se a algoritmos aglomerativos (ou, ascendentes) que atuam a partir dos n objetos iniciais, encarados como n grupos distintos, formando novos grupos por aglutinação sucessiva de grupos formados anteriormente; ou, então a algoritmos divisivos (ou, descendentes) que atuam a partir de um grupo inicial, formando novos grupos por divisão sucessiva de grupos anteriores até chegar a n grupos singulares de um só objeto (Everitt and Hothorn, 2011). Esta estrutura hierárquica proveniente de um dos algoritmos referidos representa-se, usualmente, por um gráfico de duas dimensões designado por dendrograma, que mostra como os sucessivos grupos se vão formando ao longo do processo hierárquico.

De entre os métodos de agrupamento hierárquicos, existem muitos critérios para definir a dissemelhança entre dois grupos, sendo que cada um deles dará origem a diferentes agrupamentos. Assim, a configuração dos dados deve ser examinada com cuidado. Entre estes critérios, encontram-se duas

medidas inter-grupo muito simples dadas por

$$d_{AB} = \min \{d_{im} : i \in A, m \in B\} \quad (3.5)$$

e

$$d_{AB} = \max \{d_{im} : i \in A, m \in B\}. \quad (3.6)$$

onde d_{AB} é a dissemelhança entre dois grupos A e B , designadas por método da ligação simples e método da ligação completa, respetivamente. Uma outra possibilidade para medir a dissemelhança inter-grupo é o método da ligação média em que a dissemelhança entre dois grupos é dada pela média das dissemelhanças entre todos os pares de objetos,

$$d_{AB} = \frac{\sum_{i=1}^{n_A} \sum_{m=1}^{n_B} d_{im}}{n_A n_B}, \quad (3.7)$$

onde n_A e n_B representam o número de objetos em cada grupo. Existe ainda um outro método muito utilizado designado por método de *Ward* (ou, método da variância mínima) (Reis, 2001). O algoritmo subjacente a esta técnica inicia-se com o cálculo das médias das variáveis para cada grupo. De seguida, calcula-se o quadrado da distância euclidiana entre estas médias e os valores das variáveis para cada indivíduo, somando posteriormente as distâncias para todos os indivíduos. Por fim, otimiza-se a variância mínima dentro dos grupos.

A questão que se coloca agora é qual o número de grupos a reter. Neste sentido, é alvo de interesse a comparação entre a matriz de dissemelhanças inicial e a matriz de dissemelhanças dos derivados do dendrograma. O método de comparação mais comum baseia-se no cálculo da correlação de *Pearson*, designada por correlação *cofonética* (Anderson, 2003). A magnitude desta correlação deve ser muito próxima de 1, de forma a representar uma solução de alta qualidade. No entanto, a escolha de uma medida de dissemelhança e de um método de agrupamento deve ser alvo de bom senso, uma vez que grande parte dos fenómenos em estudo apresentam uma variabilidade muito elevada.

3.1.2 Análise em componentes principais

As primeiras referências à análise em componentes principais (ACP) surgiram através de Pearson e Hotelling (Jolliffe, 2002). Esta metodologia para além de ser considerada adequada quando pretendemos obter novas variáveis não correlacionadas, a partir de um conjunto de variáveis correlacionadas, também é uma técnica apropriada para a redução da dimensionalidade de um conjunto de dados. Para tal, usa-se o conhecimento adquirido através da variabilidade dos dados, eliminando a variabilidade espúria ou sistemática, que confundem a análise estatística e, consequentemente, a sua interpretação.

Do ponto de vista geométrico, o objetivo da ACP é a identificação de um novo sistema de eixos ortogonais, em que as coordenadas das observações relativamente aos novos eixos fornecem os

valores das novas variáveis, designadas por componentes principais (Morrison, 2005). Desta forma, pretende-se encontrar as novas variáveis Y_1, Y_2, \dots, Y_p que sejam não correlacionadas, cujas variâncias decresçam da primeira à última variável. Neste tipo de abordagem, cada variável $Y_j, j = 1, \dots, p$, é obtida através de uma combinação linear das variáveis originais X_1, X_2, \dots, X_p de tal forma que

$$Y_j = a_{1j}X_1 + a_{2j}X_2 + \dots + a_{pj}X_p = \mathbf{a}_j^T \mathbf{x}, \quad (3.8)$$

onde $\mathbf{a}_j^T = [a_{1j}, a_{2j}, \dots, a_{pj}]$ é um vetor de constantes tal que $\mathbf{a}_j^T \mathbf{a}_j = 1$, fixando desta forma a escala das componentes principais (Chatfield and Collins, 1980).

De acordo com Jolliffe (2002), a primeira componente principal, Y_1 , é determinada pela escolha de \mathbf{a}_1 de maneira que esta componente tenha a maior variância possível, sujeita à restrição supracitada. Uma vez que $\text{Var}[Y_1] = \text{Var}[\mathbf{a}_1^T \mathbf{x}] = \mathbf{a}_1^T \Sigma \mathbf{a}_1$, $\mathbf{a}_1^T \Sigma \mathbf{a}_1$ é a função sujeita à restrição $\mathbf{a}_1^T \mathbf{a}_1 = 1$. O procedimento usual para maximizar uma função de várias variáveis sujeita a uma condição é o método dos multiplicadores de *Lagrange*, que dá como solução as raízes do sistema homogéneo $(\Sigma - \lambda \mathbf{I}) \times \mathbf{a}_1 = 0$ (Anderson, 2003). Para que este sistema tenha solução, para além da solução trivial, $\Sigma - \lambda \mathbf{I}$ deve ser uma matriz singular. Portanto, λ deve ser tal que $|\Sigma - \lambda \mathbf{I}| = 0$, ou seja, λ será valor próprio de Σ . Mas, uma vez que a matriz covariância é simétrica e definida positiva, Σ tem p valores próprios distintos. Como o objetivo é maximizar esta variância, então escolhe-se λ como sendo o maior valor próprio de Σ , designando-se por λ_1 , e \mathbf{a}_1 o correspondente vetor próprio associado de norma 1.

A segunda componente principal, Y_2 , é obtida como uma extensão da aquisição de Y_1 . Para além da restrição $\mathbf{a}_2^T \mathbf{a}_2 = 1$, tem-se ainda que Y_2 deve ser não correlacionada com Y_1 . Ora, $\text{Cov}[Y_2, Y_1] = \text{Cov}[\mathbf{a}_2^T \mathbf{x}, \mathbf{a}_1^T \mathbf{x}] = \mathbb{E}[\mathbf{a}_2^T (\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{a}_1] = \mathbf{a}_2^T \Sigma \mathbf{a}_1$, uma vez que se impõe média nula às várias variáveis. Portanto, $\mathbf{a}_2^T \Sigma \mathbf{a}_1 = 0$ o que é equivalente a $\mathbf{a}_2^T \mathbf{a}_1 = 0$, isto é, \mathbf{a}_1 e \mathbf{a}_2 devem ser ortogonais. Uma vez mais, usa-se o método dos multiplicadores de *Lagrange*, de forma a maximizarmos a variância de Y_2 sujeita às duas restrições definidas atrás, levando a resolver o sistema homogéneo $(\Sigma - \lambda \mathbf{I}) \times \mathbf{a}_2 = 0$, cuja solução λ é o segundo maior valor próprio, λ_2 , e \mathbf{a}_2 o correspondente vetor próprio associado de norma 1.

Continuando com este tipo de argumento, tem-se que os valores da j -ésima componente principal são definidos pelo vetor próprio de norma 1 correspondente ao j -ésimo maior valor próprio da matriz Σ .

Denotando a matriz $p \times p$ dos vetores próprios da matriz Σ por $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_p]$ e por \mathbf{y} o vetor de ordem p constituído pelas componentes principais, onde $\mathbf{y}^T = [Y_1, Y_2, \dots, Y_p]$, tem-se que

$$\mathbf{y} = \mathbf{A}^T \mathbf{x}. \quad (3.9)$$

A matriz de covariância de \mathbf{y} é dada por

$$\mathbf{\Lambda} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_p \end{pmatrix}, \quad (3.10)$$

que é uma matriz diagonal, uma vez as componentes principais foram escolhidas de forma a serem não correlacionadas. Portanto, o valor próprio λ_j de $\mathbf{\Sigma}$ pode ser interpretado como a respetiva variância da componente principal Y_j .

Sabe-se que $\text{Var} [\mathbf{y}] = \mathbf{A}^T \mathbf{\Sigma} \mathbf{A}$, pelo que $\mathbf{\Lambda} = \mathbf{A}^T \mathbf{\Sigma} \mathbf{A}$, tratando-se de uma relação muito importante uma vez que faz a ligação entre as componentes principais e a matriz de covariância de \mathbf{x} . Note-se que $\sum_{j=1}^p \text{Var} [Y_j] = \sum_{j=1}^p \lambda_j = \text{tr}(\mathbf{\Lambda}) = \text{tr}(\mathbf{A}^T \mathbf{\Sigma} \mathbf{A}) = \text{tr}(\mathbf{\Sigma} \mathbf{A} \mathbf{A}^T) = \text{tr}(\mathbf{\Sigma}) = \sum_{j=1}^p \text{Var} [X_j]$, ou seja, a soma das variâncias das p componentes principais é igual à soma das variâncias das p variáveis originais (Everitt and Dunn, 2010).

Desta forma, a componente principal Y_j explica $\frac{\lambda_j}{\sum_{k=1}^p \lambda_k} \times 100\%$ da variabilidade total dos dados,

sendo que as primeiras $l \leq p$ componentes principais explicarão $\frac{\sum_{r=1}^l \lambda_r}{\sum_{k=1}^p \lambda_k} \times 100\%$. Uma vez que a ACP não é invariante a mudanças multiplicativas diferenciadas nas escalas das variáveis, deve-se aplicar esta técnica sob a matriz de correlação \mathbf{P} em vez da matriz de covariância $\mathbf{\Sigma}$. Pelo que, a componente principal Y_j explicará $\frac{\lambda_j}{p} \times 100\%$ da variabilidade total, pois $\text{tr}(\mathbf{P}) = p$.

Assim, uma das formas de selecionar o número de componentes principais a reter é identificar aquelas que permitem explicar entre 70% e 90% da variabilidade total dos dados (Everitt and Dunn, 2010).

A equação (3.9) relaciona o vetor aleatório observado \mathbf{x} com as componentes principais \mathbf{y} de tal forma que \mathbf{y} apresenta valor médio não nulo. A transformação que se faz, neste caso, é

$$\mathbf{y} = \mathbf{A}^T (\mathbf{x} - \bar{\mathbf{x}}) \quad (3.11)$$

que consiste numa translação seguida de uma rotação ortogonal (Chatfield and Collins, 1980).

Usando a relação (3.11), na observação \mathbf{x}_i do i -ésimo indivíduo, tem-se

$$\mathbf{y}_i = \mathbf{A}^T (\mathbf{x}_i - \bar{\mathbf{x}}), \quad (3.12)$$

onde \mathbf{y}_i é designado por *score* do i -ésimo indivíduo.

3.1.3 Análise fatorial

Em muitos estudos não é possível medir diretamente as variáveis de interesse, designadas por *variáveis latentes*, *fatores comuns* ou apenas *fatores*. A ideia será avaliar outras variáveis mensuráveis, com o intuito de quantificar a variável de interesse (Morrison, 2005).

A análise fatorial (AF) ramifica-se em duas vertentes distintas: a análise fatorial exploratória (AFE), que é usada na investigação das relações entre as variáveis observáveis e as variáveis latentes sem fazer nenhuma suposição acerca destas relações; e a análise fatorial confirmatória (AFC) que é utilizada para testar se um modelo com os vários fatores, formulado à priori, fornece um ajuste adequado para as covariâncias ou correlações entre as variáveis observáveis (Everitt and Dunn, 2010).

3.1.3.1 Análise fatorial exploratória

Em geral, o primeiro passo da AFE consiste no estudo das relações entre as variáveis observáveis, utilizando o coeficiente de correlação como medida de associação entre cada par de variáveis. A aplicação da AFE permitir-nos-á identificar um número relativamente pequeno de fatores comuns às variáveis observáveis que podem ser usados para identificar relacionamentos entre um conjunto de muitas variáveis inter-relacionadas entre si. A AFE assenta num modelo de regressão múltipla cujas variáveis resposta são as variáveis observáveis e as variáveis explicativas são os fatores (Everitt and Hothorn, 2011). Todavia, a estimação dos coeficientes de regressão é muito menos direta do que numa regressão. Neste contexto, contém-se p variáveis aleatórias X_1, X_2, \dots, X_p com vetor valor médio μ e matriz de covariância Σ . Assumindo que a característica da matriz de covariância é máxima e após a standardização das variáveis observadas, o objetivo é explicar a estrutura de covariância destas variáveis.

Assim, assume-se que existem $q < p$ fatores subjacentes às variáveis observáveis, f_1, f_2, \dots, f_q , e que cada variável observada X_j é uma função linear destes fatores mais um resíduo, ou seja,

$$X_j = \lambda_{j1}f_1 + \lambda_{j2}f_2 + \dots + \lambda_{jr}f_r + \dots + \lambda_{jq}f_q + \varepsilon_j, \quad (3.13)$$

onde $j = 1, \dots, p$; λ_{jr} representa o *loading* da j -ésima variável no r -ésimo fator; e, ε_j designa o resíduo associado a esta função linear, designado por *fator específico*. Note-se que os *loadings* mostram o quanto as variáveis observáveis dependem do j -ésimo fator.

Este modelo, tal como a regressão múltipla, estará sujeito a várias hipóteses (Jobson, 1992):

- os fatores comuns apresentam valor médio nulo, $\mathbb{E}[f_r] = 0$, e variância unitária, $\text{Var}[f_r] = 1$, onde $r = 1, \dots, q$;
- os fatores comuns são independentes uns dos outros, sendo que esta hipótese pode ser relaxada, pois pode-se aplicar a AFE com uma rotação não ortogonal da matriz dos fatores;

- os fatores específicos são independentes entre si e independentes dos fatores comuns;
- a variância do fator específico de índice j , $\text{Var} [\varepsilon_j]$, é representada por ψ_j .

Desta forma, pode-se construir a matriz que contém os *loadings* subjacentes a todas as variáveis observáveis, designada por Λ . O modelo da AFE, representado na equação (3.13), pode ser reescrito na forma matricial por

$$\mathbf{x} = \Lambda \mathbf{f} + \boldsymbol{\varepsilon}, \quad (3.14)$$

onde

$$\Lambda = \begin{pmatrix} \lambda_{11} & \lambda_{12} & \cdots & \lambda_{1q} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{p1} & \lambda_{p2} & \cdots & \lambda_{pq} \end{pmatrix}, \mathbf{f} = \begin{pmatrix} f_1 \\ \vdots \\ f_q \end{pmatrix} \text{ e } \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_p \end{pmatrix}.$$

Neste contexto, Λ denota uma matriz $p \times q$, não podendo ser confundida com a matriz diagonal dos valores próprios apresentada com o mesmo símbolo na redução canónica de Σ na metodologia da ACP.

Tendo em consideração a equação (3.13) e usando a hipótese de que os fatores são independentes com variância unitária, tem-se

$$\text{Var} [X_j] = \lambda_{j1}^2 + \lambda_{j2}^2 + \cdots + \lambda_{jq}^2 + \text{Var} [\varepsilon_j] \quad (3.15)$$

$$= \sum_{k=1}^q \lambda_{jk}^2 + \psi_j = h_j^2 + \psi_j \quad (3.16)$$

onde a parte relativa à variância explicada pelos fatores comuns, $h_j^2 = \sum_{k=1}^q \lambda_{jk}^2$, denomina-se por comunalidade da j -ésima variável.

Da equação (3.13), pode-se ainda observar que $\text{Cov} [X_k, X_l] = \sum_{r=1}^q \lambda_{kr} \lambda_{lr}$, para $k \neq l$. Assim, a matriz de covariância de \mathbf{X} pode ser obtida por

$$\Sigma = \Lambda \Lambda^T + \Psi, \quad (3.17)$$

onde Ψ representa uma matriz diagonal cujas entradas são as variâncias específicas, $\psi_1, \psi_2, \dots, \psi_p$. O termo geral ψ_j representa a variabilidade específica que não é partilhada com as outras variáveis (Anderson, 2003). Ora, se a decomposição da equação (3.17) se verificar, é correto assumir um modelo contendo q fatores. Na prática, a matriz de covariância Σ é desconhecida e será estimada pela matriz de covariância amostral S (ou pela matriz de correlação amostral \mathbf{R}). Desta forma, precisa-se de encontrar estimativas para Λ e Ψ de tal forma que $S \approx \Lambda \Lambda^T + \Psi$.

Atualmente, existem muitos métodos para proceder à estimação dos parâmetros do modelo da AFE (Bartholomew and Knott, 1999), entre os quais:

- *fator principal*, que tem por base a estimação das comunalidades, tendo como base o cálculo

dos valores e vetores próprios de \mathbf{S} (ou de \mathbf{R}) tal como na técnica de ACP. No entanto, efetua-se o cálculo sobre a matriz de covariância reduzida, \mathbf{S}^* , definida por $\mathbf{S}^* = \mathbf{S} - \mathbf{\Psi}$;

- *máxima verosimilhança*, cujo teste de ajustamento ao modelo com q fatores é baseado numa distância F entre a matriz \mathbf{S} e a matriz $\hat{\mathbf{S}}$ predita, onde $F = \log(|\mathbf{\Lambda}\mathbf{\Lambda}^T + \mathbf{\Psi}|) + \text{tr}(\mathbf{S}|\mathbf{\Lambda}\mathbf{\Lambda}^T + \mathbf{\Psi}|^{-1}) - \log(|\mathbf{S}|) - p$, que toma o valor zero se $\mathbf{S} = \mathbf{\Lambda}\mathbf{\Lambda}^T + \mathbf{\Psi}$ e valores superiores a zero, caso contrário. Através da minimização de F obtém-se estimativas para os *loadings* dos fatores comuns e para as variâncias específicas. Uma desvantagem deste método é a exigência da multinormalidade dos dados;
- *mínimos quadrados*, que tem implementado a minimização da matriz de correlação residual diagonal, $\mathbf{\Psi} = \mathbf{\Sigma} - \mathbf{\Lambda}\mathbf{\Lambda}^T$, ajustando os valores próprios da matriz de correlação inicial. Os resultados deste método muito similares ao método de máxima verosimilhança.

Uma vez que não existe uma solução única para a matriz dos *loadings* dos fatores, considera-se o modelo equivalente com *loadings* dados por $\mathbf{\Lambda}^* = \mathbf{\Lambda}\mathbf{T}$, para uma matriz particular e ortogonal \mathbf{T} - rotação dos fatores. Esta rotação é um processo segundo o qual uma solução pode ser passível de interpretação sem alterar as propriedades matemáticas subjacentes. Nesta análise, pode-se considerar dois tipos de rotação: a rotação ortogonal (que são métodos restritos à condição da preservação da ortogonalidade dos fatores); e a rotação oblíqua (que são métodos que permitem obter fatores correlacionados). Uma das vantagens da rotação ortogonal reside no facto de os *loadings* representarem as correlações entre os fatores e as variáveis originais, o que não acontece quando a rotação é oblíqua, devido à correlação entre os fatores (Chatfield and Collins, 1980).

Em muitas situações torna-se necessário o cálculo dos *scores* dos fatores para cada indivíduo da amostra (coincidindo, neste estudo, com o número de localizações observadas). A equação que define o modelo da AF, equação (3.13), não é invertível. As variáveis originais encontram-se expressas em termos dos fatores, mas para calcular os *scores* deve ter-se essa relação na direção oposta. Uma vez obtidas as estimativas para $\mathbf{\Lambda}$ e $\mathbf{\Psi}$, é usado o método dos mínimos quadrados ponderados com o intuito de se obter estimativas para os *scores* dos fatores, $\hat{\mathbf{f}}$, dados por $\hat{\mathbf{f}} = (\hat{\mathbf{\Lambda}}^T \hat{\mathbf{\Psi}}^{-1} \hat{\mathbf{\Lambda}})^{-1} \hat{\mathbf{\Lambda}}^T \hat{\mathbf{\Psi}}^{-1} \mathbf{x}$ (Anderson, 2003).

3.1.3.2 Análise fatorial confirmatória

A AFC é um método estatístico incluído num conjunto de técnicas de modelação de variáveis latentes, conhecido como Modelação de Equações Estruturais (MEE), que tem por objetivo testar se um modelo complexo, contendo os vários fatores, fornece um ajuste adequado para as covariâncias ou correlações entre as variáveis mensuráveis.

A metodologia MEE contém um número de parâmetros que precisam ser estimados através da matriz de covariância ou de correlação das variáveis observáveis. A estimação destes valores opera através da minimização de uma função de discrepância que indica a magnitude das diferenças entre

os elementos de S e de $\Sigma(\theta)$, matriz de covariância implícita pelo modelo ajustado, onde θ é um vetor que contém todos os parâmetros do modelo a estimar.

Existem muitas possibilidades para definir a função de discrepância, como por exemplo a função de discrepância dos mínimos quadrados ordinários, FMQ, dada por

$$\text{FMQ}(S, \Sigma(\theta)) = \sum_{k < l} \sum_l [s_{kl} - \sigma_{kl}(\theta)]^2, \quad (3.18)$$

onde s_{kl} e $\sigma_{kl}(\theta)$ são elementos de S e $\Sigma(\theta)$, respetivamente (Everitt and Hothorn, 2011). No entanto, esta função apresenta algumas desvantagens, como por exemplo o não ser independente da escala das variáveis mensuráveis (Everitt, 1984).

A função de discrepância mais utilizada para a estimação dos vários parâmetros numa AFC é a de máxima verosimilhança, FMV, sob o pressuposto da multinormalidade dos dados observados, dada por

$$\text{FMV}(S, \Sigma(\theta)) = \log |\Sigma(\theta)| - \log |S| + \text{tr}(S\Sigma(\theta)^{-1}) - p, \quad (3.19)$$

sobre a qual se tem de aplicar algoritmos iterativos de forma a minimizar esta função (Timm, 2002).

Uma vez estimados os parâmetros do modelo e identificada a matriz de covariância implícita pelo modelo considerado, o passo seguinte é avaliar o quão bem a matriz de covariância do modelo predito se ajusta à matriz de covariância das variáveis observáveis. Uma medida global que avalia este ajuste é fornecida pela estatística da razão de verosimilhança dada por $X^2 = (n - 1)\text{FMV}_{\min}$, onde n é a dimensão da amostra e FMV_{\min} é o valor mínimo da função de discrepância de máxima verosimilhança patente na equação (3.19). Se o número da amostra for suficientemente grande, a estatística X^2 fornece um teste, cuja hipótese nula é: a matriz de covariância populacional das variáveis observáveis é igual à matriz de covariância subjacente ao modelo ajustado. Sob a veracidade da hipótese nula, X^2 segue uma distribuição χ^2 com $\nu = \frac{1}{2}p(p + 1) - v$ graus de liberdade, onde v é o número de parâmetros presentes no modelo (Everitt and Dunn, 2010).

A estatística da razão de verosimilhança é, muitas vezes, a única medida de qualidade indicada pelo modelo ajustado. No entanto esta tem um uso prático limitado porque quando se tem um número muito elevado de observações ter-se-á a rejeição da hipótese nula. Uma forma satisfatória de utilizar este teste é a comparação de uma série de modelos aninhados, onde uma grande diferença na estatística de teste de dois modelos, em comparação com a diferença dos graus de liberdade destes, indica que os parâmetros adicionais num dos modelos proporciona uma verdadeira melhoria do ajuste. Adicionalmente a este teste, deve-se usar outros procedimentos (Everitt and Dunn, 2010), nomeadamente

- inspecionando visualmente as covariâncias residuais, que devem ser bastante menores quando comparadas com as covariâncias ou correlações observadas;
- examinando os erros padrão dos parâmetros e as correlações entre estas estimativas. Se estas correlações forem elevadas poderá significar que o modelo é mal ajustado aos dados.

Existem diversos indicadores da qualidade do ajustamento. Entre eles encontra-se um indicador denominado por RMSR (*root mean square residual*), que é a raiz quadrada da média das diferenças quadráticas entre os elementos de \mathbf{S} e $\hat{\Sigma}(\theta)$. Um valor de RMSR inferior a 0.05 é geralmente considerado como uma indicação de um ajustamento razoável.

RMSEA (*root mean square error of approximation*) é outro indicador que indica a discrepância entre os resíduos previstos e dos observados, caso o modelo seja estimado para a população. Valores inferiores a 0.06 indicam um bom ajuste.

GFI (*goodness of fit index*) é outro indicador da qualidade do ajustamento. Este índice baseia-se na razão da soma do quadrado das distâncias entre a matriz de covariância observada e a a matriz de covariância reproduzida pelo modelo. Pelo método de estimação de máxima verosimilhança, o valor GFI é dado por

$$\text{GFI} = 1 - \frac{\text{tr} \left[\left(\mathbf{S} \hat{\Sigma}(\theta)^{-1} - \mathbf{I} \right) \left(\mathbf{S} \hat{\Sigma}(\theta)^{-1} - \mathbf{I} \right) \right]}{\text{tr} \left(\mathbf{S} \hat{\Sigma}(\theta)^{-1} \mathbf{S} \hat{\Sigma}(\theta)^{-1} \right)}, \quad (3.20)$$

que pode variar entre zero (sem ajustamento) e um (ajustamento perfeito). Na prática, valores acima de 0.9 sugerem um nível aceitável do ajustamento (Everitt and Hothorn, 2011).

3.2 Métodos de análise geoestatística

Os métodos estatísticos para dados espaciais têm sido uma área de intensa investigação nos últimos trinta anos e, ainda hoje, existem muitos tópicos por investigar. Muitos dos métodos utilizados nesta área partem do seguinte princípio: pontos próximos no espaço tendem a ter valores mais similares do que pontos mais afastados - Lei de Tobler, Primeira Lei da Geografia (Tobler, 1970). A natureza desse espaço permite identificar três dos principais processos espaciais: os processos reticulados, os processos pontuais e os processos contínuos (referidos como geoestatística) (Carvalho and Natário, 2008).

Os modelos de geoestatística trabalham com observações de uma variável aleatória de interesse em distintas localizações espaciais, medindo a sua relação com o objetivo de representar a noção intuitiva da existência de uma correlação entre dados situados numa proximidade espacial. Tipicamente, esta correlação tende a diminuir com o aumento da distância. O conhecimento destas relações é particularmente relevante de um ponto de vista prático, uma vez que permite fazer predição dos valores para as localizações em que não foram efetuadas medições.

Ao contrário dos métodos de geoestatística, os métodos determinísticos têm por base critérios estritamente geométricos que podem não ser coerentes com a estrutura espacial subjacente ao processo em estudo, não fornecendo medidas de incerteza relacionadas com o fenómeno, com as observações e com o método de inferência (Soares, 2000; Goovaerts, 1997). Existem diversos estudos que demonstram que os métodos geoestatísticos, nomeadamente os métodos de *kriging*, apresentam melhores resultados em comparação com os modelos determinísticos (Dien and Comrie, 2002; Lloyd

and Atkinson, 2004).

A aplicação da geoestatística terá então como objetivo a obtenção do conhecimento da dispersão espacial das grandezas que definem a quantidade e a qualidade dos fenómenos naturais. Estes fenómenos espaciais, nas Ciências da Terra e do Ambiente, têm algumas especificidades próprias que os distinguem de outros, que podem ser objeto da aplicação de modelos matemáticos e físicos, uma vez que existe um acentuado grau de incerteza ligado ao seu conhecimento (Soares, 2000).

Os dados analisados consistem em medidas de uma variável de interesse num conjunto finito de locais, $\{\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_\alpha, \dots, \mathbf{s}_n\}$, pertencentes a um certo domínio de área não nula, $A \subseteq \mathbb{R}^d$ (geralmente, estas localizações encontram-se num espaço bidimensional, ou seja, $d = 2$). Neste âmbito, dispõe-se de um conjunto de dados

$$\begin{pmatrix} \mathbf{s}_1 & y_1 \\ \mathbf{s}_2 & y_2 \\ \vdots & \vdots \\ \mathbf{s}_\alpha & y_\alpha \\ \vdots & \vdots \\ \mathbf{s}_n & y_n \end{pmatrix} \quad (3.21)$$

onde cada observação $y_\alpha = y(\mathbf{s}_\alpha)$, $\alpha = 1, 2, \dots, n$, é considerada como uma realização de uma variável aleatória $Y_\alpha = Y(\mathbf{s}_\alpha)$, cuja distribuição é governada por um processo estocástico subjacente espacialmente contínuo $\{Y(\mathbf{s}), \mathbf{s} \in A\}$, representado por Y , por simplificação. Os valores Y_α são identicamente distribuídos mas não independentes, uma vez que poderão estar espacialmente correlacionados (Wackernagel, 2003).

Os locais de observação isto é, o desenho amostral, do processo serão selecionados de acordo com planos de amostragem, nomeadamente de acordo com o plano não preferencial considerado determinístico ou estocasticamente independente do processo Y (Diggle and Ribeiro, 2007). Este processo diz-se, ainda, não uniforme se a intensidade da amostragem variar sistematicamente ao longo da região de estudo. Os principais objetivos científicos da geoestatística são

- estimação, referindo-se a inferências sobre os parâmetros do modelo estocástico subjacente aos dados;
- predição, referindo-se a inferências sobre a realização do processo não observado Y .

3.2.1 Hipóteses requeridas

Os processos espaciais subjacentes terão de verificar diversas hipóteses, nomeadamente que a média e a variância, representadas respetivamente por $\text{IE}[Y(\mathbf{s})] = \mu(\mathbf{s})$ e $\text{Var}[Y(\mathbf{s})] = \sigma^2(\mathbf{s})$, são finitas.

Ter-se-á, ainda, de admitir a hipótese da **estacionaridade** do processo, que é descrita detalhadamente por Cressie (1993). Geralmente, apenas é necessário que os dois primeiros momentos sejam

invariantes a translações, ou seja, que o processo seja estacionário de segunda ordem ou fracamente estacionário,

$$\mathbb{E} [Y(\mathbf{s})] = \mu, \mathbf{s} \in A \quad (3.22)$$

e

$$\text{Cov} [Y(\mathbf{s}_i), Y(\mathbf{s}_j)] = C(\mathbf{s}_i - \mathbf{s}_j), \quad (3.23)$$

onde $i, j = 1, \dots, n$ e $\mathbf{s}_i, \mathbf{s}_j \in A$ (ou seja, se a covariância depende apenas do vetor diferença entre \mathbf{s}_i e \mathbf{s}_j). A função $C(\cdot)$ é usualmente designada por função covariância ou covariograma.

Esta definição implica que o processo tenha variância constante, $\sigma^2(\mathbf{s}) = C(\mathbf{0})$, e, ainda, que o covariograma seja simétrico, $C(\mathbf{h}) = C(-\mathbf{h})$. Se $C(\mathbf{0}) \neq 0$, pode definir-se a propriedade (3.23) à custa da função de correlação ou correlograma $\rho(\cdot)$

$$\text{Corr} [Y(\mathbf{s}_i), Y(\mathbf{s}_j)] = \frac{\text{Cov} [Y(\mathbf{s}_i), Y(\mathbf{s}_j)]}{C(\mathbf{0})} = \rho(\mathbf{s}_i - \mathbf{s}_j), \quad (3.24)$$

para a qual se verifica $\rho(\mathbf{0}) = 1$ e $\rho(\mathbf{h}) = \rho(-\mathbf{h})$.

No entanto, basta existir a presença de uma tendência espacial para que se torne inadequado supor que a média seja constante; ou ainda, ocorrer casos em que não seja possível obter-se a covariância como função de um vetor diferença. Assim, por vezes assume-se um tipo de estacionaridade mais fraco. Supondo que a média e a variância dos incrementos do processo, $Y(\mathbf{s}_i) - Y(\mathbf{s}_j)$, existem e são independentes da localização \mathbf{s} , um processo diz-se intrinsecamente estacionário se

$$\mathbb{E} [Y(\mathbf{s}_i) - Y(\mathbf{s}_j)] = 0, \quad (3.25)$$

e

$$\text{Var} [Y(\mathbf{s}_i) - Y(\mathbf{s}_j)] = \mathbb{E} \left[\{Y(\mathbf{s}_i) - Y(\mathbf{s}_j)\}^2 \right] = 2\gamma(\mathbf{s}_i - \mathbf{s}_j) \quad (3.26)$$

onde $\mathbf{s}_i, \mathbf{s}_j \in A$. A função $\gamma(\cdot)$ designa-se por semivariograma (ou, variograma por simplificação).

Tendo em conta que $\gamma(\mathbf{s}_i - \mathbf{s}_j) = \frac{1}{2} \{ \text{Var} [Y(\mathbf{s}_i)] + \text{Var} [Y(\mathbf{s}_j)] - 2\text{Cov} [Y(\mathbf{s}_i), Y(\mathbf{s}_j)] \}$, o variograma de um processo estocástico espacial Y , é dado por

$$\gamma(\mathbf{s}_i - \mathbf{s}_j) = \sigma^2 [1 - \rho(\mathbf{s}_i - \mathbf{s}_j)]. \quad (3.27)$$

Segundo Carvalho and Natário (2008), o processo espacial pode ser decomposto em vários sub-processos: $Y(\mathbf{s}) = \mu(\mathbf{s}) + W(\mathbf{s}) + \epsilon(\mathbf{s})$, cada um deles refletindo o comportamento do fenómeno numa determinada escala,

- uma primeira componente determinística, $\mu(\cdot)$, designada por tendência espacial, reflete o comportamento médio do processo e traduz a variabilidade em larga escala;
- uma segunda componente, $W(\cdot)$, de valor médio nulo, fracamente estacionária e contínua em

média quadrática, modela a ligação espacial entre as observações e traduz a variabilidade em pequena e micro escala¹;

- uma terceira componente, $\epsilon(\cdot)$, pretende representar os erros de medida e é modelada por um ruído branco, independente da componente anterior, sob o pressuposto que, se fossem feitas várias observações do processo numa certa localização, estas flutuariam em torno do correspondente valor do processo.

Desta forma, os variogramas de processos espaciais estacionários de segunda ordem são, genericamente, representados pela expressão

$$\gamma(\mathbf{s}_i - \mathbf{s}_j) = \tau^2 + \sigma^2 [1 - \rho(\mathbf{s}_i - \mathbf{s}_j)]. \quad (3.28)$$

Normalmente, o variograma tende para um valor constante à medida que a distância $\mathbf{s}_i - \mathbf{s}_j$ aumenta, sendo este valor designado por *sill* o qual corresponde à variância total do processo Y , $\sigma^2 + \tau^2$. Quando o variograma apresenta este valor (sendo possível especificá-lo), significa que existe uma distância a partir da qual a correlação entre as observações é zero, a esta distância dá-se o nome de *range* correspondendo a um parâmetro identificado por ϕ . O parâmetro τ^2 , variância do ruído, designa-se por *nugget effect*.

Um outro requisito requerido ao processo espacial é a **isotropia**, definindo-se à custa do variograma. A isotropia implica que a covariância entre os valores de $Y(\mathbf{s})$ dependa apenas de uma distância entre as respetivas localizações, ou seja, se quaisquer que sejam $\mathbf{s}_i, \mathbf{s}_j \in A$, $C(\mathbf{s}_i - \mathbf{s}_j) = C(\|\mathbf{s}_i - \mathbf{s}_j\|)$ (Stein, 1999). No caso de verificar-se a isotropia do processo, os valores do *sill* dos variogramas direcionais têm valor igual. A falta da hipótese de isotropia num processo, pode ser causada por comportamentos diferentes do processo segundo as várias direções do espaço, sendo a mais frequente designada por anisotropia geométrica, que tem por base a variação direcional do *range* do variograma. Um processo que seja simultaneamente intrinsecamente estacionário e isotrópico chama-se homogéneo e um processo que não é isotrópico diz-se anisotrópico.

3.2.2 Variograma empírico

Supondo que o processo espacial subjacente aos dados é, no mínimo, intrinsecamente estacionário e isotrópico, a associação entre as diversas observações pode ser modelada pelo variograma omnidirecional. A modelação do variograma pode ser comparada com a modelação paramétrica da distribuição teórica de uma variável aleatória. Desta forma, num primeiro passo é criado um estimador do variograma baseado na amostra, que se designa por variograma empírico. Este estimador é uma ferramenta muito útil para a análise exploratória dos dados, podendo servir para seleccionar de entre

¹Por vezes, esta parcela da decomposição é, ainda, decomposta em duas componentes, $W(\mathbf{s}) = W_1(\mathbf{s}) + W_2(\mathbf{s})$, estacionárias de segunda ordem e independentes, onde $W_1(\cdot)$ e $W_2(\cdot)$ explicam a variação em pequena e micro escala, respetivamente. No entanto, a menos que existam observações repetidas nas mesmas localizações, é impossível distinguir o efeito de $W_2(\cdot)$ e de $\epsilon(\cdot)$.

a família paramétrica de variogramas, o modelo de correlação espacial adequado para os dados em questão. É através de métodos de regressão não paramétrica como por exemplo o método de *Kernel* ou de regressão por *splines*, que surgem estimadores alisados do variograma. No entanto, existem desvantagens na sua aplicação uma vez que as observações do variograma empírico são dependentes (Diggle and Ribeiro, 2007).

Em 1962, Matheron apresenta, usando o método dos momentos, um estimador do variograma, designado por estimador clássico, dado por

$$\hat{\gamma}(u) = \frac{1}{2 |N(u)|} \sum_{N(u)} [Y(\mathbf{s}_i) - Y(\mathbf{s}_j)]^2, \quad (3.29)$$

onde $N(u) = \{(\mathbf{s}_i, \mathbf{s}_j) : \|\mathbf{s}_i - \mathbf{s}_j\| = u, u \in \mathbb{R}\}$ e $|N(u)|$ é o respetivo cardinal (Goovaerts, 1997). Este estimador, apesar de ser centrado quando o processo é intrinsecamente estacionário, apresenta alguns inconvenientes tal como o de poder ser afetado por valores anómalos devido ao termo ao quadrado que aparece no somatório da fórmula (3.29). Existe um outro estimador mais robusto a *outliers*, sendo necessário a verificação da normalidade do processo (Cressie, 1993).

O estimador do variograma apresentado na fórmula (3.29) não satisfaz a propriedade de ser uma função semidefinida negativa, ou seja, não verifica a condição $\sum_{i=1}^n \sum_{j=1}^n a_i a_j \gamma(\mathbf{s}_i - \mathbf{s}_j) \leq 0$, para um qualquer conjunto finito de localizações espaciais $\{\mathbf{s}_1, \dots, \mathbf{s}_n\}$ e um qualquer conjunto de números reais $\{a_1, \dots, a_n\}$, tais que $\sum_{i=1}^n a_i = 1$. Portanto, este estimador não deve ser usado num contexto de inferência e predição. Assim, usam-se modelos de variograma válidos pertencentes às famílias paramétricas.

Famílias paramétricas de variogramas

Alguns modelos teóricos isotrópicos verificam as propriedades dos variogramas, nomeadamente a propriedade supracitada. Estas curvas fazem parte de uma das famílias paramétricas do tipo

$$P = \{\gamma : \gamma(\cdot) = \gamma(\cdot; \theta), \theta \in \Theta\},$$

sendo que $\gamma(\cdot; \theta)$ representa uma função condicionalmente definida negativa dependendo dos valores dados pelo vetor de parâmetros Θ , destacando-se as famílias *Matérn* e *Esférica* (Diggle and Ribeiro, 2007).

- *Família Matérn*

Esta família foi apresentada por Matérn em 1960, correspondendo a um conjunto de modelos com aspetos diversificados, uma vez que este modelo depende de mais de um parâmetro. Esta é dada por

$$\gamma(u; \theta) = \theta_0 + \theta_1 (1 - \rho(u; \phi = \theta_2; \kappa = \theta_3)) \quad (3.30)$$

onde $\theta = (\theta_0, \theta_1, \theta_2, \theta_3)^T$, $\rho(u; \phi; \kappa) = \frac{1}{2^{\kappa-1}\Gamma(\kappa)} \left(\frac{u}{\phi}\right)^\kappa K_\kappa\left(\frac{u}{\phi}\right)$, com $K_\kappa(\cdot)$ denotando a função de *Bessel* modificada de segunda ordem. Os parâmetros θ_0 e θ_1 identificam, respetivamente, τ^2 e σ^2 .

Caso $k = 0.5$ a função de correlação *Mátern* $\rho(u; \phi = \theta_2; \kappa = \theta_3)$, patente na equação (3.30), reduz-se à Exponencial dada por $\rho(u; \phi) = \exp\left(-\frac{u}{\phi}\right)$; e, se $k \rightarrow \infty$ obtém-se a função de correlação Gaussiana, onde $\rho(u; \phi) \rightarrow \exp\left(-\left(\frac{u}{\phi}\right)^2\right)$.

- **Família Esférica**

Esta família é uma das mais adotadas devido ao facto de fornecer interpretações diretas dos seus parâmetros, nomeadamente para ϕ , e para $\theta_0 + \theta_1$ que corresponde ao *sill*. Este modelo é dado por

$$\gamma(u; \theta) = \begin{cases} \theta_0 + \theta_1 \left(\frac{3u}{2\theta_2} - \frac{1}{2} \left(\frac{u}{\theta_2} \right)^3 \right), & 0 < u \leq \theta_2 \\ \theta_0 + \theta_1, & u > \theta_2 \end{cases} \quad (3.31)$$

onde $\theta = (\theta_0, \theta_1, \theta_2)^T$.

3.2.3 Estimação

Existe um particular interesse na estimação dos parâmetros envolvidos nas estruturas do valor médio e da associação espacial do modelo adotado para o processo estocástico que se supõe gerar as observações. Relativamente à estrutura de covariância dos dados, pretende-se estimar a variância σ^2 de Y , a variância do erro de medição τ^2 e o valor do *range* ϕ e outros parâmetros que estejam contidos no modelo teórico de variograma escolhido.

Tendência espacial

Uma qualquer variação da média designa-se por tendência espacial. A forma mais simples de estimar a tendência espacial de forma a que o processo seja estacionário é permitir que a resposta média, $\mu(\mathbf{s})$, possa depender da localização. Na prática, pode-se escolher diretamente um modelo que seja função de \mathbf{s} , que frequentemente é obtido através de um modelo de regressão polinomial, usando somatórios e produtos cruzados das coordenadas cartesianas de \mathbf{s} como variáveis explicativas. Modelos deste tipo são designados por modelos de superfície de tendência (Carvalho and Natário, 2008).

De acordo com Diggle and Ribeiro (2007), as superfícies de tendência linear ou quadrática podem fornecer descrições empíricas úteis e simples das inexplicáveis tendências espaciais, contudo as superfícies de grau superior devem ser evitadas uma vez as tendências complicadas são descritas muito melhor por meio da componente estocástica do modelo.

Um tipo mais interessante de tendência espacial surge quando a função média pode ser modelada

utilizando covariáveis espacialmente referenciadas, ou seja,

$$Y(\mathbf{s}) = \mu(\mathbf{s}) + \delta(\mathbf{s}) = \beta_0 + \sum_{k=1}^p \beta_k X_k(\mathbf{s}) + \delta(\mathbf{s}) \quad (3.32)$$

em que X_k , $k = 1, \dots, p$, são variáveis explicativas que podem ou não depender da localização e $\delta(\mathbf{s}) = W(\mathbf{s}) + \epsilon(\mathbf{s})$ é um processo de média nula e variância finita que poderá ou não ter correlação espacial. Modelos deste tipo são mais interessantes do que os modelos de superfície de tendência uma vez que estes procuram explicar, ao invés de apenas descrever, a variação espacial da variável resposta.

Uma vez estimada a tendência espacial, o procedimento da estimação dos parâmetros da estrutura de correlação é sobre a componente $\delta(\mathbf{s}) = Y(\mathbf{s}) - \mu(\mathbf{s})$.

Isotropia

A continuidade espacial de um dado recurso ou fenómeno natural nem sempre é igual nas várias direções do espaço. O principal objetivo da modelação de estruturas anisotrópicas é reduzir as diferentes estruturas de continuidade das diferentes direções a um único modelo, ou seja, é transformar estas estruturas de modo a serem equivalentes a uma estrutura isotrópica. O modelo de anisotropia mais comum é o da anisotropia geométrica. Este tipo de anisotropia caracteriza-se por os variogramas apresentarem o mesmo modelo com o mesmo valor do *sill* em todas as direções, mas com valores de *range* diferentes.

Esta ocorre quando uma estrutura de covariância estacionária é transformada por um diferencial de alongamento e de rotação dos eixos de coordenadas. Daí, a anisotropia geométrica ser definida por dois parâmetros adicionais. Algebricamente, um modelo com anisotropia geométrica com coordenadas espaciais $\mathbf{s} = (s_1, s_2)$ pode ser convertido a um modelo estacionário com coordenadas $\mathbf{s}' = (s'_1, s'_2)$, pela transformação

$$(s'_1, s'_2) = (s_1, s_2) \begin{bmatrix} \cos(\psi_A) & -\sin(\psi_A) \\ \sin(\psi_A) & \cos(\psi_A) \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \psi_R^{-1} \end{bmatrix}, \quad (3.33)$$

onde ψ_A e $\psi_R > 1$ são designados, respetivamente por *ângulo* e *rácio* da anisotropia (Diggle and Ribeiro, 2007). A direção ao longo da qual a correlação diminui mais lentamente com o aumento da distância é chamado eixo principal. Note-se que a anisotropia geométrica não pode descrever características direcionais locais de uma superfície espacial, apenas globais. Por outro lado, a presença de características direcionais locais numa realização de um processo espacial, não implica necessariamente que o processo subjacente seja anisotrópico.

Variograma

Em geoestatística, o variograma não é apenas utilizado para fins exploratórios, mas também para a estimação dos parâmetros formais. Para estimar estes parâmetros, pode-se utilizar o método dos mínimos quadrados ou o método de máxima verosimilhança. Ao apresentar estes métodos, supõe-se que está identificada a tendência espacial e também que se está sob a posse do variograma empírico da componente $\delta(\mathbf{s})$, cujo objetivo será estimar o vetor de parâmetros θ .

Método dos mínimos quadrados O método dos mínimos quadrados é o mais conhecido, podendo ser escrito em notação matricial

$$\min_{\theta} \{(\hat{\gamma} - \gamma_{\theta})^T W^{-1} (\hat{\gamma} - \gamma_{\theta})\}, \quad (3.34)$$

onde $\hat{\gamma}$ identifica um estimador empírico do variograma e γ_{θ} o respetivo modelo paramétrico válido, cuja forma é conhecida exceto o vetor de parâmetros θ (Carvalho and Natário, 2008). A matriz W é uma matriz de pesos onde

- se W corresponder à matriz identidade, então estar-se-á perante o critério de mínimos quadrados ordinários. Neste caso, este método pode ser reescrito por

$$\min_{\theta} \left\{ \sum_{i=1}^L (\hat{\gamma}(u_i) - \gamma_{\theta}(u_i))^2 \right\}, \quad (3.35)$$

onde L representa o número total de *lags* (isto é, distâncias) usadas na estimação. Este método apresenta como principal vantagem a sua simplicidade de aplicação, não envolvendo informação sobre a variância e a covariância da distribuição do estimador empírico, nem do número de pares de pontos envolvidos na estimação de $lag(u)$;

- se $W = V$, onde V é a matriz de variância, cujos elementos são dados por $v_{ij} = \text{Cov} [\hat{\gamma}(u_i), \hat{\gamma}(u_j)]$, então estar-se-á perante o critério de mínimos quadrados generalizados. Este método apresenta como principal desvantagem o ser demasiado complexo, dependendo da função desconhecida $\hat{\gamma}(\cdot)$;
- se $W = D$, onde D é uma matriz diagonal, cujos elementos desta são dados por $d_{ii} = \text{Var} [\hat{\gamma}(u_i)]$, então estar-se-á perante o critério de mínimos quadrados pesados. Este é um caso em que é balanceada a eficiência do critério de mínimos quadrados generalizados e a simplicidade do critério de mínimos quadrados ordinários.

Método da máxima verosimilhança Pelo facto das observações não admitirem a hipótese de independência, criando algumas dificuldades na aplicação de métodos de regressão não linear, é usual fazer hipóteses adicionais sobre a distribuição do modelo em questão, usando-se métodos de máxima

verosimilhança para estimar todos os parâmetros do modelo sem recorrer a nenhum tipo de variograma estimado (Diggle and Ribeiro, 2007). Dentro do contexto da geoestatística, a implementação deste método apenas é direto e simplificado quando os dados são gerados por um modelo gaussiano.

Suponha-se que $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)^T \sim N(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}(\theta))$, onde \mathbf{X} é a matriz de ordem $n \times p$ com as observações de p covariáveis, $\boldsymbol{\beta}$ é o vetor de coeficientes de regressão e $\boldsymbol{\Sigma}(\theta)$ é a matriz de covariância que pode ser escrita como

$$\boldsymbol{\Sigma}(\theta) = \tau^2 \mathbf{I} + \sigma^2 \mathbf{R}(\phi) \quad (3.36)$$

onde $\mathbf{R}(\phi)$ corresponde à matriz de correlação (dependente de um parâmetro escalar ou vetorial ϕ), cujos elementos são dados por $r_{ij} = \rho(\|\mathbf{s}_i - \mathbf{s}_j\|; \phi)$, $i, j = 1, \dots, n$, e \mathbf{I} corresponde à matriz identidade de ordem n . Consequentemente, os parâmetros do modelo podem ser estimados maximizando a função de log-verosimilhança dada por

$$L(\boldsymbol{\beta}, \tau^2, \sigma^2, \phi) = -\frac{1}{2} \{n \log(2\pi) + \log(|\tau^2 \mathbf{I} + \sigma^2 \mathbf{R}(\phi)|) + (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\tau^2 \mathbf{I} + \sigma^2 \mathbf{R}(\phi))^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\}. \quad (3.37)$$

Alguns *softwares*, como a *package* geoR, resolvem o problema de otimização de forma iterativa, sendo possível especificar valores iniciais para os parâmetros.

3.2.4 Validação cruzada

A validação cruzada é uma forma de verificar qual o modelo teórico mais adequado aos dados em análise. Assim, deve-se especificar um modelo de variograma e uma vizinhança de pesquisa, comparando-se posteriormente, os valores reais com os preditos, sendo a diferença entre eles designada por resíduo da validação cruzada. A ideia subjacente a este método é utilizar os resíduos para verificar a performance do modelo. Note-se que esta metodologia não prova que o modelo do variograma está correto, apenas prova que ele não é grosseiramente incorreto.

A ideia fundamental desta técnica é estimar Y em cada localização amostrada \mathbf{s}_i à custa dos dados vizinhos $Y(\mathbf{s}_j)$, com $j \neq i$, eliminando da base de dados original a observação $Y(\mathbf{s}_i)$. Desta forma, em cada ponto amostrado \mathbf{s}_i obtém-se uma predição $\hat{Y}(\mathbf{s}_i)$ e a respetiva variância σ_i^2 , sendo os valores de predição influenciados pelo variograma estimado $\hat{\gamma}$. A qualidade do ajustamento pode ser medida através do erro médio (EM)

$$EM = \frac{1}{n} \sum_{i=1}^n Y(\mathbf{s}_i) - \hat{Y}(\mathbf{s}_i), \quad (3.38)$$

que deve ser aproximadamente igual a zero, ou através da média standardizada do erro quadrático (MEEQ)

$$MEEQ = \frac{1}{n} \sum_{i=1}^n \frac{(Y(\mathbf{s}_i) - \hat{Y}(\mathbf{s}_i))^2}{\sigma_i^2}, \quad (3.39)$$

que ser aproximadamente igual a um.

De acordo com Soares (2000), em relação ao modelo do variograma, não se pretende que este seja um estimador ergódico de um segundo momento de uma função aleatória $Y(\mathbf{s})$. Pretende-se, acima de tudo, que ele reflita, em termos médios, os principais padrões de continuidade espacial e da variabilidade do conjunto de amostras $Y(\mathbf{s}_i)$ e que proporcione o conhecimento (muitas vezes, não mensurável) do fenómeno em estudo. Devido a este facto, na maioria das situações, é preferível condicionar o ajustamento do variograma a um dado modelo ao conhecimento do processo em causa, do que condicioná-lo a estatísticas resultantes de ajustamentos de mínimos quadrados.

3.2.5 Predição

Conhecido o modelo teórico que fora ajustado ao variograma do processo em causa e havendo dependência espacial entre os valores da amostra, pretende-se descobrir, dado um conjunto de n observações do processo Y nas localizações $\{\mathbf{s}_i : i = 1, 2, \dots, n\}$, qual o valor assumido pela variável $Y(\mathbf{s}_0)$ numa localização \mathbf{s}_0 , não amostrada.

Segundo Goovaerts (1997), todos os estimadores de *kriging* são derivados de um mesmo estimador de regressão linear básico $\hat{Y}(\mathbf{s}_0)$ definido por

$$\hat{Y}(\mathbf{s}_0) - \mu(\mathbf{s}_0) = \sum_{i=1}^n \lambda_i [Y(\mathbf{s}_i) - \mu(\mathbf{s}_i)] \quad (3.40)$$

onde os coeficientes λ_i , $i = 1, \dots, n$, são designados por pesos de predição e são tais que $\sum_{i=1}^n \lambda_i = 1$. Esta combinação linear é considerada ótima se o respetivo estimador for centrado e apresentar variância mínima. Desta forma, com o objetivo de determinar os pesos de predição, λ_i , minimiza-se a variância do estimador

$$\sigma_K^2 = \text{Var} \left[\hat{Y}(\mathbf{s}_0) - \mu(\mathbf{s}_0) \right] \quad (3.41)$$

sob a restrição $\mathbb{E} \left[\hat{Y}(\mathbf{s}_0) - \mu(\mathbf{s}_0) \right] = 0$, sendo que apenas os dados contidos num raio de influência devem ser considerados.

Neste tópico, a estimação de um variograma tem um papel fundamental uma vez que é usado para encontrar a solução ótima para os valores dos pesos λ_i . O método designa-se por *kriging*, podendo ser aplicado a um processo estacionário de segunda ordem ou intrínseco apresentando média nula. Assim, o variograma que entra como *input* no método de *kriging* deve representar a componente residual da variável, $\delta(\mathbf{s}) = Y(\mathbf{s}) - \mu(\mathbf{s})$. Os detalhes dos vários métodos de *kriging* são descritos detalhadamente por Goovaerts (1997) e Soares (2000).

O método mais simples de predição denomina-se por ***kriging simples***. Este supõe que a componente da tendência é uma média constante e conhecida, isto é, $\mu(\mathbf{s}_0) = \mu$. Assim sendo, a média da população é utilizada para cada estimação local, em conjunto com os pontos vizinhos estabelecidos como necessários. Este método tem por objetivo a minimização do erro quadrático médio, con-

siderando as estimativas dos parâmetros do modelo verdadeiras. Desta forma, a estimação do erro $\hat{Y}_{KS}(\mathbf{s}_0) - Y(\mathbf{s}_0)$ é obtida minimizando a quantidade $\mathbb{E} \left[\left(\hat{Y}_{KS}(\mathbf{s}_0) - Y(\mathbf{s}_0) \right)^2 \right]$.

Outro método de predição denomina-se por **kriging ordinário**, onde neste caso, a componente da tendência é uma média constante mas desconhecida. As médias locais não são necessariamente próximas da média da população utilizando-se, portanto, apenas os pontos vizinhos para a estimação de um ponto não amostrado. O sistema de *kriging*, neste método, pode ser escrito à custa da função do variograma $\gamma(u)$, onde $u = \mathbf{s}_i - \mathbf{s}_j$, tendo em conta que $\gamma(u) = C(0) - C(u)$. Este estimador é centrado (ou seja, a média dos valores preditos será igual à média dos valores reais) e apresenta variância mínima (implicando que quanto mais dispersa é a nuvem dos valores preditos versus valores reais, mais impreciso será o estimador). No entanto, este estimador tem como hipótese a estacionaridade da média, ou seja, a média é desconhecida mas constante ao longo de toda a região espacial A , que na maioria das situações não se verifica.

Para considerar o caso em que a média do processo é variável, torna-se necessário adaptar a fórmula de *kriging* para $\hat{Y}(\mathbf{s}_0)$, substituindo a constante μ por uma tendência espacial $\mu(\mathbf{s})$ variável. Este procedimento é associado ao método de **kriging universal** ou **kriging com tendência externa**.

Capítulo 4

Biomonitorização da Deposição Atmosférica dos Elementos Químicos

Neste estudo são utilizadas diversas metodologias de análise estatística com o objetivo de compreender qual o comportamento estatístico dos dados, bem como compreender as relações existentes entre os diversos elementos químicos amostrados. O conhecimento estatístico dos dados, através da análise exploratória, constitui uma etapa fundamental da análise que deve preceder a modelação geoestatística (Soares, 2000).

Desta forma, neste capítulo são apresentados os resultados de uma análise descritiva univariada, determinando-se vários parâmetros estatísticos descritivos, sendo posteriormente, efetuado um estudo geoestatístico onde são aplicados modelos de *kriging* ordinário e universal (consoante os elementos químicos), com o intuito de construir mapas de predição para os diversos elementos. Posteriormente, é apresentado um estudo multivariado das quatro campanhas onde são utilizadas as várias metodologias de análise multivariada referidas na secção 3.1, com o intuito de detetar os locais mais poluídos de Portugal continental, bem como identificar as principais relações entre os vários elementos químicos.

As várias metodologias de análise estatística foram realizadas utilizando o *software* R (R Team, 2010), versão 2.15.2 (Everitt and Hothorn, 2011; Diggle and Ribeiro, 2007; Crawley, 2007).

4.1 Caracterização dos elementos químicos - análise univariada

4.1.1 Análise exploratória dos elementos químicos

Com o intuito de descobrir as principais propriedades e compreender o comportamento de cada um dos elementos químicos em estudo, procedeu-se a uma análise estatística univariada através do cálculo das principais características amostrais das concentrações dos vários elementos nos musgos, como média, mínimo, máximo, etc., apresentadas na Tabela 4.1.

Analisando as estatísticas descritivas das concentrações dos elementos químicos nos musgos associadas a cada uma das campanhas, verifica-se que grande parte dos elementos apresentam

um elevado valor para o coeficiente de variação¹, implicando a existência de uma elevada dispersão destas variáveis relativamente à média. Verifica-se, ainda, que as várias variáveis apresentam um valor elevado para o coeficiente de assimetria², indicando a predominância de valores baixos e uma assimetria positiva da distribuição subjacente. As concentrações dos elementos apresentam, também, uma grande variabilidade.

É de salientar que em algumas campanhas foram amostradas outras espécies de musgos, como se verifica pela Figura 2.2, pelo facto de não existir nenhuma das espécies principais no local ou por falha de identificação da espécie na amostragem (Sérgio et al., 1993). A utilização de diferentes espécies de musgo pode ser considerada como uma eventual razão para o aumento da variabilidade global dos resultados (na campanha de 2006 não existe este problema). Por esta razão, em estudos anteriores, foi testada a necessidade de aplicação de fatores de correção entre as várias espécies de musgo, a partir da comparação de várias amostras das mesmas em 13 locais. Os resultados da análise de variância para este estudo evidenciaram que não existiam diferenças significativas entre os dois principais musgos ($p - \text{valor} > 0.05$) (Figueira et al., 2002). Um outro fator que poderá ter contribuído para a elevada variabilidade final dos resultados é o facto de, nas várias campanhas, as amostragens realizarem-se num longo período de tempo. Consequentemente, verifica-se alterações das concentrações dos elementos nos briófitos, influenciados, por exemplo pela precipitação e pelo vento; sendo que esta variabilidade terá sido introduzida aleatoriamente, pela razão de que a amostragem não seguiu uma sequência temporal rígida.

De forma a compreender melhor a evolução destes parâmetros estatísticos para os diversos elementos, elaboraram-se diagramas de extremos e quartis para cada variável ao longo das quatro campanhas, presentes no Anexo III, representando-se apenas os *outliers* severos. Verifica-se que, à exceção do elemento químico Cu, as concentrações dos restantes elementos parecem sofrer um aumento da campanha de 1992 para a de 1997. Note-se que durante estes anos não existia uma preocupação tão cuidada sobre o ambiente como nos dias de hoje em Portugal. Da campanha de 1997 para a de 2002 evidencia-se uma diminuição da concentração dos elementos, assim como da campanha de 2002 para a de 2006 (sendo, no entanto, um decrescimento menor), à exceção dos elementos As, Hg e V que parecem manter-se constantes ao longo destas campanhas. Esta diminuição está muito vincada no elemento químico Pb, Figura 13(i).

Verifica-se também a presença de *outliers* severos em todos os elementos químicos, estando os valores mais elevados patentes nos elementos Cr e Ni (campanha de 1997), Cu (campanha de 2002) e V (campanha de 2006). Do ponto de vista estatístico, estes *outliers* deveriam ser removidos, pois podem prejudicar os procedimentos de análise estatística, no entanto estes valores têm significado geoquímico.

¹O coeficiente de variação é interpretado como a variabilidade dos dados em relação à média. Quanto menor for o valor deste coeficiente, mais homogêneo é o conjunto de dados.

²O coeficiente de assimetria possibilita analisar a simetria de uma dada distribuição. Uma distribuição é dita simétrica quando o seu valor é, aproximadamente, zero.

Tabela 4.1: Parâmetros estatísticos das concentrações dos vários elementos químicos nos musgos observadas em cada campanha.

	Ano	n ^a	Média	Mediana	D.p. ^b	C.v. ^c	Mínimo	Máximo	Q25 ^d	Q75 ^e	C.a. ^f
(mg/kg)											
As	2002	151	0.78	0.33	1.40	1.79	0.01	9.71	0.01	0.95	3.94
	2006	98	1.24	0.57	2.45	1.98	0.03	19.32	0.27	1.18	5.28
Cd	1992	173	0.22	0.09	0.32	1.49	0.04	1.90	0.04	0.24	2.96
	1997	177	0.78	0.70	0.52	0.67	0.04	3.53	0.44	0.95	2.00
	2002	151	0.50	0.41	0.48	0.96	0.04	4.10	0.25	0.62	4.67
	2006	98	0.04	0.01	0.13	3.28	0.00	1.26	0.00	0.04	8.28
Cr	1992	173	2.23	1.33	2.72	1.22	0.10	18.00	0.80	2.40	3.22
	1997	177	3.86	1.95	9.01	2.33	0.04	107.26	1.00	3.78	9.02
	2002	151	1.66	1.07	2.13	1.28	0.27	19.55	0.67	1.75	4.93
	2006	98	0.86	0.42	1.48	1.72	0.07	11.52	0.22	0.98	4.94
Cu	1992	173	9.34	8.00	7.09	0.76	1.20	52.00	5.20	11.00	3.25
	1997	177	8.91	7.37	7.10	0.80	0.40	52.01	4.88	10.72	3.22
	2002	151	7.34	6.13	6.73	0.92	2.81	80.36	4.84	7.79	8.61
	2006	98	4.47	3.67	3.49	0.78	1.70	27.05	2.75	5.03	4.20
Fe	1992	173	1163.32	820.00	1127.29	0.97	130.00	7260.00	468.00	1398.00	2.68
	1997	177	1940.13	1207.75	2030.90	1.05	82.98	11193.61	663.85	2381.35	2.23
	2002	151	833.35	562.53	821.19	0.99	139.84	5113.34	390.97	911.61	2.95
	2006	98	415.07	305.88	351.17	0.85	80.88	1852.26	178.09	464.81	1.93
Hg	2002	151	0.09	0.04	0.22	2.56	0.00	1.74	0.03	0.07	6.33
	2006	98	0.07	0.06	0.06	0.78	0.01	0.43	0.04	0.09	2.99
Mn	1992	173	158.39	123.00	140.30	0.89	16.00	970.00	72.00	198.00	2.43
	1997	177	188.26	152.27	154.21	0.82	4.03	948.76	86.07	246.05	1.77
	2002	151	147.41	125.38	98.31	0.67	23.14	503.11	75.41	196.07	0.99
	2006	98	118.59	84.93	116.27	0.98	2.24	625.79	34.60	162.67	1.80
Ni	1992	173	2.47	1.80	2.89	1.17	0.15	25.00	1.00	2.80	4.26
	1997	177	13.55	9.49	23.24	1.72	0.83	254.46	5.96	14.50	8.17
	2002	151	1.97	1.21	3.27	1.65	0.01	26.83	0.56	1.98	5.23
Pb	1992	173	17.46	13.00	16.97	0.97	0.50	172.00	9.00	21.00	5.19
	1997	177	21.75	16.66	22.86	1.05	1.71	191.17	10.12	26.86	4.59
	2002	151	5.40	3.11	10.24	1.90	0.68	109.93	2.20	4.92	7.70
	2006	98	1.93	1.23	1.79	0.93	0.10	9.38	0.77	2.41	2.01
Zn	1992	173	39.48	29.00	52.56	1.33	4.30	660.00	22.00	44.00	9.65
	1997	177	52.43	40.41	59.69	1.14	12.97	709.79	29.02	58.28	7.98
	2002	151	38.33	27.73	39.74	1.04	12.98	332.58	21.97	40.14	5.40
	2006	98	30.62	22.84	28.16	0.92	9.32	213.08	17.08	31.12	3.82
V	2002	151	3.62	2.72	3.39	0.94	0.26	20.09	1.65	4.16	2.52
	2006	98	17.32	2.77	105.22	6.08	0.23	1039.33	1.92	4.22	9.36

^aNúmero de observações na respetiva campanha

^bDesvio-padrão

^cCoeficiente de variação

^dPrimeiro quartil

^eTerceiro quartil

^fCoeficiente de assimetria

4.1.2 Análise e modelação da continuidade espacial dos elementos químicos

O estudo geoestatístico do fenómeno em causa envolve, essencialmente, o conhecimento das suas medidas de continuidade espacial com o intuito de caracterizar a sua distribuição e/ou variabilidade espacial. Note-se que para cada $\mathbf{s} \in \mathbb{R}^2$, $Y(\mathbf{s})$ é uma variável aleatória que apenas apresenta uma única realização do fenómeno, uma vez que trabalhar com realizações espaciais, tipicamente, implica a impossibilidade da repetição de várias observações numa dada localização. No entanto, para ser possível a inferência a partir de amostras de dimensão unitária, deve-se impor condições sobre o comportamento que está na base do processo (desconhecido) que produz estas variáveis aleatórias $\{Y(\mathbf{s})\}_{\mathbf{s} \in \mathbb{R}^2}$: a estacionaridade e a isotropia.

Após a verificação das hipóteses descritas, segue-se a descrição da estrutura espacial da família de variáveis aleatórias através do cálculo do variograma empírico, que é obtido através da média do quadrado da diferença entre todos os pares de pontos, em diferentes classes de distância. A existência de *outliers* é particularmente importante aquando da modelação do variograma. Optou-se por excluir estes valores, representados a vermelho nos diagramas de extremos e quartis (Anexo III), durante a estimação dos parâmetros do modelo teórico do variograma, sendo novamente incluídos na matriz de dados na fase da predição de valores para locais não amostrados. Após o ajuste de um modelo teórico ao variograma empírico, realizou-se a interpolação da variável a toda a região em estudo, através dos métodos *kriging* ordinário ou *kriging* universal. No caso do *kriging* universal recorreu-se à regressão linear simples, considerando-se como eventuais variáveis explicativas a latitude e a longitude.

A apresentação mais detalhada sobre a aplicação do procedimento supracitado é feita com o elemento químico As, fazendo-se uma breve explicação do procedimento estatístico do elemento em causa, apresentando simultaneamente, os mapas de predição e respetiva interpretação. Para os restantes elementos a apresentação das Tabelas contendo as estimativas dos vários parâmetros e dos gráficos dos variogramas empíricos e teóricos é feita no Anexo IV. Para determinadas campanhas dos elementos químicos verifica-se, através dos respetivos variogramas, a não existência de continuidade espacial. Nestes casos, torna-se pouco interessante a modelação do variograma empírico e, como consequência, a aplicação do método *kriging* para a obtenção dos mapas de predição.

Arsénio

A primeira etapa da análise geoestatística é a estimação dos parâmetros envolvidos na estrutura do valor médio, contida no modelo adotado para o processo estocástico que se supõe gerar as observações. Estes parâmetros estão contidos na Tabela 4.2, com o respetivo desvio-padrão. Verifica-se que, para ambas as campanhas, a componente da tendência espacial $\mu(\mathbf{s})$ tem de ser modelada através de um modelo de superfície de tendência linear, onde neste caso $\hat{\mu}(\mathbf{s}) = \hat{\beta}_0 + \hat{\beta}_1 \text{Longitude} + \hat{\beta}_2 \text{Latitude}$, cujos parâmetros $\hat{\beta}_0$, $\hat{\beta}_1$ e $\hat{\beta}_2$ representam os valores dos coeficientes de regressão (ver Tabela 4.2).

O modelo considerado também inclui parâmetros para o ângulo e para o rácio, para o caso de o

processo apresentar anisotropia geométrica. Caso as estimativas resultantes sejam igual a zero e a um, respetivamente, verifica-se a isotropia do processo. Na campanha de 2002, o processo estocástico subjacente às concentrações de As apresenta anisotropia, podendo ser convertido a um modelo estacionário através da transformação das coordenadas, de acordo com os parâmetros: $\hat{\psi}_A = 0.29$ e $\hat{\psi}_R = 1$. Para a campanha de 2006, verifica-se a isotropia do processo.

Tabela 4.2: Estimativas dos parâmetros envolvidos nas estruturas do valor médio e da anisotropia das concentrações de As nos musgos observadas em cada campanha.

Campanha	Tendência						Anisotropia	
	$\hat{\beta}_0$	D.p.	$\hat{\beta}_1$	D.p.	$\hat{\beta}_2$	D.p.	$\hat{\psi}_A$	$\hat{\psi}_R$
2002	0.68	0.08	0.001	0.001	-0.002	0.001	0.29	1.00
2006	0.41	0.16	0.003	0.002	0.005	0.002	0.00	1.00

Uma vez estimada a componente relativa à tendência espacial e os parâmetros relativos à anisotropia, segue-se a análise da estrutura da associação espacial, cujas estimativas dos valores do variograma empírico são calculadas utilizando os resíduos: $\hat{\delta}(\mathbf{s}) = Y(\mathbf{s}) - \hat{\mu}(\mathbf{s})$, usando para tal o estimador clássico (fórmula (3.29)). Na Tabela 4.3 estão representados os valores obtidos para a construção dos variogramas empíricos, verificando-se que os variogramas, relativos às campanhas de 2002 e 2006, apresentam uma distância máxima igual a 300 Km e a 220 Km, respetivamente.

Tabela 4.3: Valores dos variogramas empíricos das concentrações de As nos musgos observadas em cada campanha, onde $|N(u)|$ representa o cardinal das diferentes classes de distâncias.

Campanha							
2002				2006			
u	$ N(u) $	$\hat{\gamma}(u)$	D.p.	u	$ N(u) $	$\hat{\gamma}(u)$	D.p.
0.00	22.00	0.26	0.42	0.00	55.00	0.35	0.53
20.00	314.00	0.57	1.93	24.44	526.00	0.56	1.15
40.00	480.00	0.61	1.85	48.89	737.00	0.57	1.18
60.00	567.00	0.70	1.92	73.33	874.00	0.61	1.19
80.00	645.00	0.87	2.40	97.78	768.00	0.64	1.16
100.00	710.00	0.83	2.18	122.22	642.00	0.60	1.11
120.00	780.00	0.86	2.02	146.67	458.00	0.63	1.11
140.00	801.00	0.93	2.38	171.11	189.00	0.58	1.03
160.00	779.00	0.96	2.53	195.56	73.00	0.67	1.14
180.00	694.00	0.85	2.20	220.00	18.00	0.68	0.98
200.00	620.00	1.02	2.66				
220.00	548.00	0.83	2.02				
240.00	520.00	0.68	1.98				
260.00	461.00	0.79	2.11				
280.00	404.00	0.85	2.22				
300.00	204.00	0.75	2.07				

Segue-se a estimação dos parâmetros desconhecidos dos modelos teóricos paramétricos que melhor se adequem aos variogramas empíricos. Foram considerados os modelos exponencial, esférico e gaussiano. Os ajustamentos dos variogramas teóricos foram realizados através do método de mínimos quadrados, uma vez que produz melhores resultados que o método de mínimos quadrados ponderados. Este método consiste em medir a qualidade do ajustamento com base na minimização do valor da soma dos quadrados das diferenças entre os valores do variograma nuvem e o do variograma teórico (soma dos quadrados dos resíduos, SQR). Para comparação dos diferentes modelos apresenta-se, também, as medidas do ME e do MEEQ, fórmulas (3.38) e (3.39). A escolha do modelo teórico foi realizado através da aproximação dos valores ME e MEEQ a zero e a um, respetivamente, e minimizando o valor do SQR, cujos valores estão apresentados na Tabela 4.4. Os modelos selecionados para ajustar aos variogramas empíricos das concentrações de As das campanhas de 2002 e 2006 foram o modelo esférico e exponencial, respetivamente.

Nos variogramas verifica-se a existência do *nugget effect*, ou seja, estes exibem uma ordenada na origem não nula que expressa a variabilidade associada aos erros de medição do fenómeno ou, ainda, a variabilidade da amostra associada a uma escala mais detalhada do que a informação disponível.

Tabela 4.4: Estimativas dos parâmetros envolvidos nas estruturas da associação espacial dos modelos adotados para as concentrações de As nos musgos observadas em cada campanha.

Campanha	Modelo	Parâmetros					
		EM	MEEQ	SQR	$\hat{\tau}^2$	$\hat{\sigma}^2$	$\hat{\phi}$
2002	Exponencial	0.006	1.41	57.52	0.24	0.64	34.60
	Esférico	0.002	1.37	53.29	0.40	0.47	113.92
	Gaussiano	0.001	1.40	53.52	0.48	0.39	57.70
2006	Exponencial	0.001	1.10	1.81	0.38	0.24	21.14
	Esférico	-0.001	1.14	2.41	0.48	0.15	91.72
	Gaussiano	-0.0001	1.15	2.91	0.50	0.12	46.13

Nas Figuras 4.1(a) e 4.1(b), apresentam-se graficamente, para as campanhas de 2002 e 2006, os variogramas amostrais das concentrações de As nos musgos observadas em cada campanha e respetivos modelos teóricos selecionados.

Obtidos os variogramas teóricos para as concentrações de As nos musgos observadas em cada campanha, o objetivo seguinte é encontrar estimativas ótimas das concentrações de As em locais onde estas não foram medidas. A interpolação dos valores de As para a totalidade do território de Portugal continental efetuou-se através do *kriging* universal, cujas superfícies preditas das concentrações de As estão presentes na Figura 4.2. Os resultados são apresentados em conjunto com os mapas da variabilidade associada a cada estimativa contidos no Anexo IV. Neste anexo são, ainda, apresentados os mapas deste elemento sem a partilha de escala.

É com base na identificação de fontes de contaminação local, que será feita a interpretação dos padrões de distribuição de cada elemento químico determinados nos musgos.

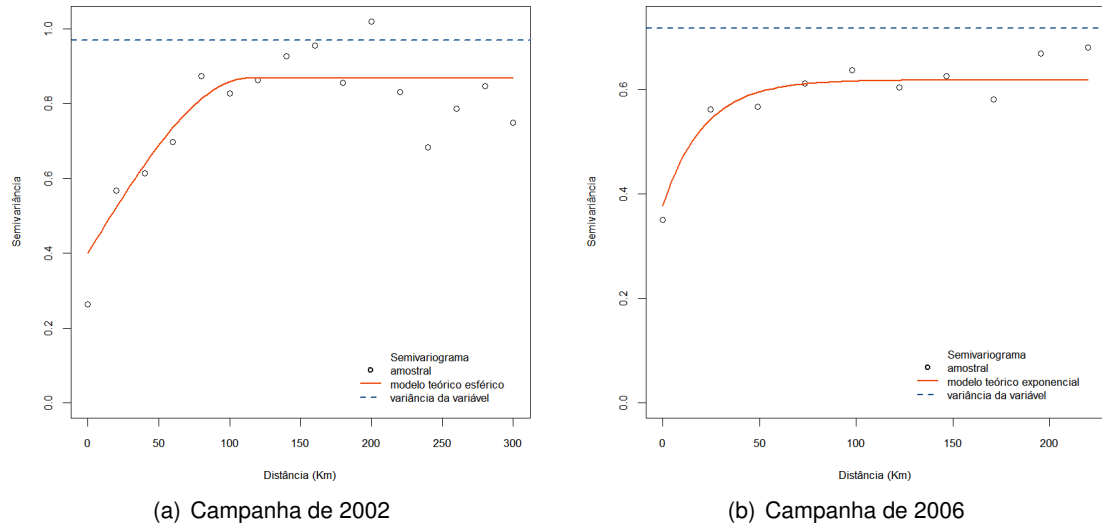


Figura 4.1: Variogramas amostrais das concentrações de As nos musgos observadas em cada campanha e respetivos modelos esférico (à esquerda) e exponencial (à direita), estimados pelo método de mínimos quadrados.

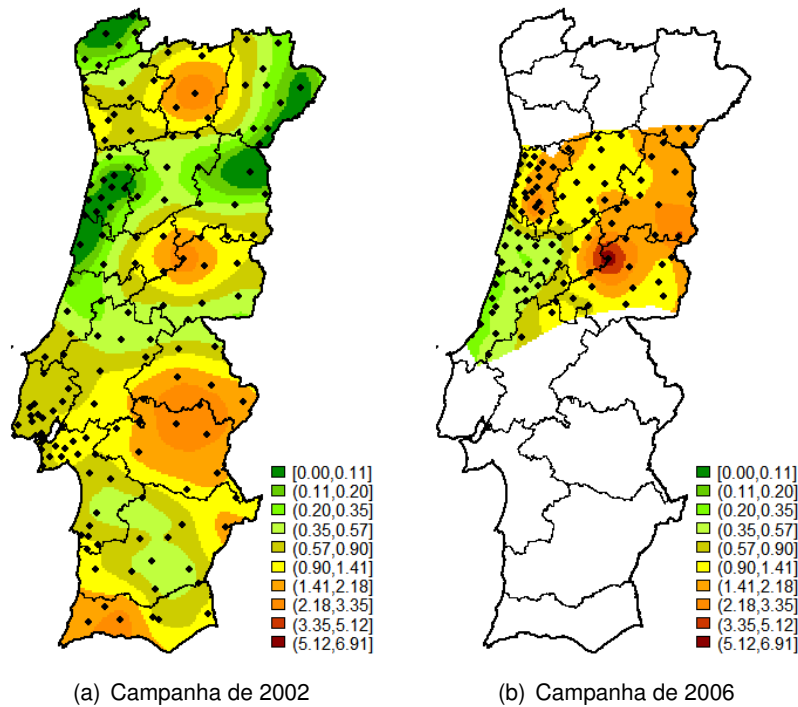


Figura 4.2: Superfícies previstas das concentrações de As nos musgos observadas em cada campanha.

Através da análise da Figura 4.2, evidencia-se, para a campanha de 2002, a influência de atividades antropogénicas nas áreas metropolitanas de Lisboa e Porto e no distrito de Sines, assim como em áreas que lhes estão associadas, sendo que as maiores concentrações de As encontram-se nos distritos de Évora/Portalegre, Castelo Branco/Coimbra e Vila Real, correspondendo às explorações de ouro e volfrâmio.

Verifica-se ainda que, ao nível da região Centro, existe um aumento das concentrações deste elemento de uma campanha para a outra, nomeadamente nos distritos de Aveiro, Viseu e Guarda, continuando a área mineira da Panasqueira a apresentar os valores mais elevados deste elemento.

Cádmio

Os processos espaciais subjacentes às concentrações de Cd nos musgos nas campanhas de 1992 e 2006 são considerados não estacionários mas isotrópicos. Como consequência, em ambos os casos, recorreu-se ao modelo de superfície de tendência linear para estimar os coeficientes de regressão relativos à componente média.

Posteriormente, os modelos exponencial e esférico foram os selecionados para modelar as concentrações de Cd nos musgos nas campanhas de 1992 e 2006, respetivamente, cujos variogramas empíricos e teóricos estão apresentados na Figura 14 (Anexo IV). As estimativas para os parâmetros dos variogramas das campanhas de 1992 e 2006 são $(\widehat{\sigma^2}, \widehat{\phi}, \widehat{\tau^2}) = (0.02, 42.40, 0.01)$ e $(\widehat{\sigma^2}, \widehat{\phi}, \widehat{\tau^2}) = (0.0004, 45.23, 0.0008)$, respetivamente.

Desta forma, procedeu-se à aplicação do método *kriging* universal, obtendo-se os mapas de predição apresentados nas Figuras 4.3(a) e 4.3(b).

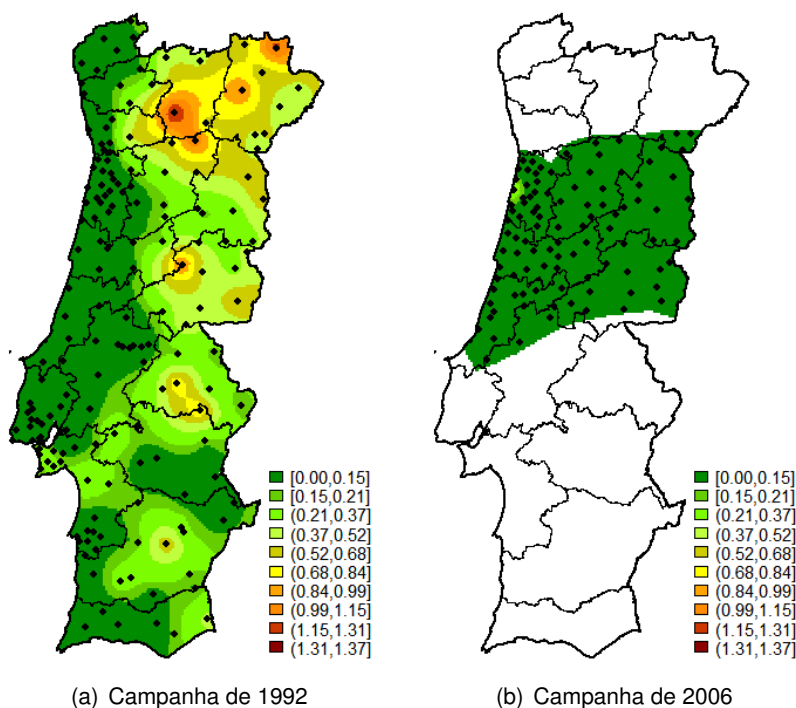


Figura 4.3: Superfícies preditas das concentrações de Cd nos musgos observadas em cada campanha.

De acordo com a Figura 4.3, verifica-se a diminuição dos valores de Cd da campanha de 1992 para a de 2006.

Na campanha de 1992 é notório o aumento dos níveis de Cd do litoral para o interior do país (nomeadamente, nos distritos de Vila Real e Bragança) cujos solos contêm fertilizantes devido à agricultura. As elevadas concentrações correspondem à ação da componente geológica e à presença de áreas mineiras (como o caso da mina de Jales, distrito de Vila Real) e à forte erosão provocada nas zonas florestais e rurais que foram afetadas pelos fogos na altura da amostragem desta campanha.

Crómio

A componente da tendência espacial, para ambas as campanhas, foi modelada através de uma superfície de tendência linear. Uma vez modelada a componente da tendência espacial, falta assumir a isotropia dos processos subjacentes às concentrações de Cr nos musgos. Observando a Tabela 4 (Anexo IV), pode-se concluir que o processo relativo à campanha de 1992 pode ser considerado isotrópico. No que diz respeito à campanha de 2002, com o objetivo de que este processo seja isotrópico, considera-se a transformação das coordenadas através dos parâmetros $(\hat{\psi}_A, \hat{\psi}_R) = (0.23, 2.66)$.

De forma a aplicarmos o método de *kriging* universal, consideraram-se o modelo esférico com parâmetros $(\hat{\sigma}^2, \hat{\phi}, \hat{\tau}^2) = (5.90, 150.48, 1.52)$, para a campanha de 1992, e o modelo esférico com parâmetros $(\hat{\sigma}^2, \hat{\phi}, \hat{\tau}^2) = (1.21, 47.01, 0.28)$, para a campanha de 2002, como sendo as funções de correlação dos variogramas teóricos.

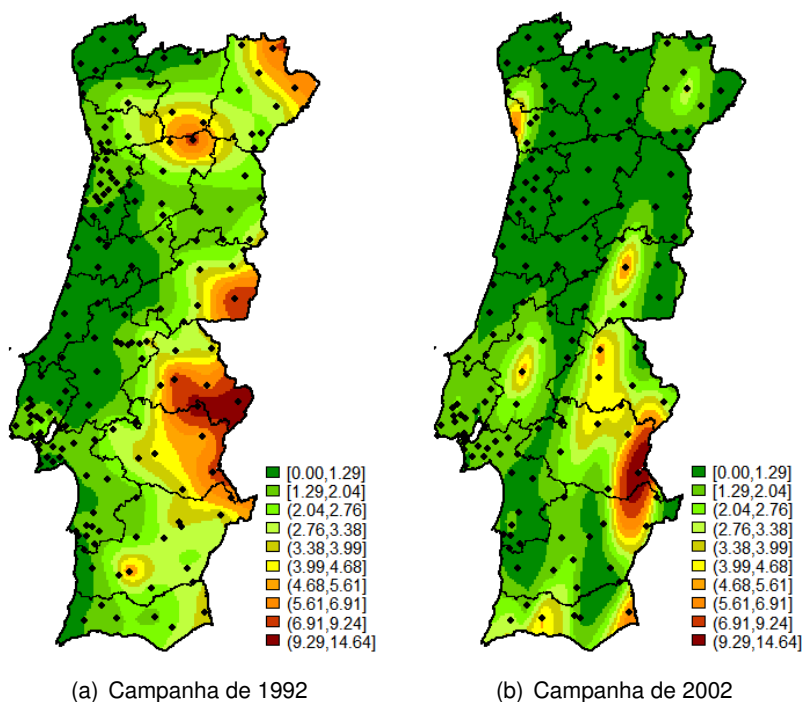


Figura 4.4: Superfícies previstas das concentrações de Cr nos musgos observadas em cada campanha.

Analisando a Figura 4.4(a), pode-se concluir que os níveis máximos nas zonas rurais podem estar ligados à contaminação dos solos por erosão, ao predomínio de rochas eruptivas incluindo as ultrabásicas, assim como metamórficas. Verifica-se, ainda, que toda a zona litoral das regiões Norte e Centro apresentam valores pouco elevados deste elemento.

Relativamente à campanha de 2002, observa-se que os valores mais elevados estão contidos no sul do país.

Verifica-se, através da Figura 4.4, que existe um decréscimo dos valores deste elemento, principalmente nas regiões Norte e Centro.

Cobre

Relativamente às concentrações de Cu, verifica-se a estacionaridade dos processos subjacentes a ambas as campanhas. Todavia, não se verifica a isotropia, sendo por isso necessário a transformação das coordenadas de acordo com os parâmetros $(\hat{\psi}_A, \hat{\psi}_R) = (0.07, 2.47)$, para a campanha de 1992, e $(\hat{\psi}_A, \hat{\psi}_R) = (0.37, 1.46)$, para a campanha de 2002.

Para a obtenção dos mapas de predição do elemento Cu, Figura 4.5, aplicou-se o método *kriging* ordinário, sendo para tal necessário a especificação das funções de correlação dos respetivos variogramas teóricos. Assim, consideraram-se, para a campanha de 1992, o modelo esférico com parâmetros $(\hat{\sigma}^2, \hat{\phi}, \hat{\tau}^2) = (9.65, 35.20, 9.65)$ e, para a campanha de 2002, o modelo exponencial com parâmetros $(\hat{\sigma}^2, \hat{\phi}, \hat{\tau}^2) = (5.61, 26.40, 2.56)$.

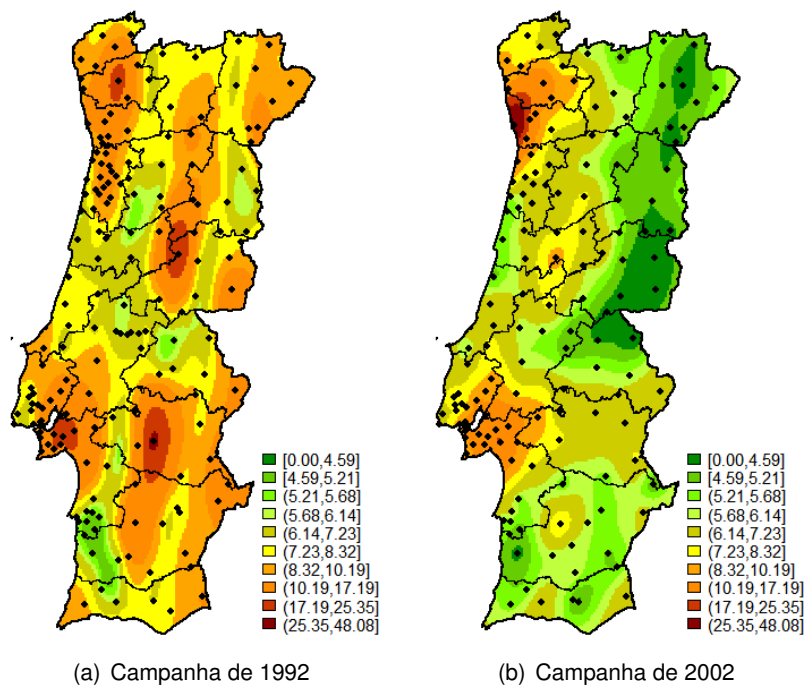


Figura 4.5: Superfícies preditas das concentrações de Cu nos musgos observadas em cada campanha.

Analisando a Figura 4.5, verifica-se que existiu um decréscimo dos valores deste elemento, nomeadamente nas zonas do interior do país, da campanha de 1992 para a de 2002.

Na campanha de 1992 observa-se valores elevados nas minas da área de Lisboa-Setúbal, sendo que estes também são notórios na zona industrial a sul do Tejo e Sado e na zona norte entre Porto e Braga, possivelmente devido às atividades industriais.

Relativamente à campanha de 2002, verifica-se concentrações elevadas nas áreas de maior atividade antropogénica como Lisboa e Porto, devido à atividade industrial (metalurgia) e emissões do tráfego automóvel. Os valores elevados presentes na zona do Douro (campanha de 1992) e na região do Minho (campanha de 1992 e 2002) podem corresponder à ação contínua de soluções ricas em Cu (fungicidas) usadas no tratamento das vinhas.

Ferro

Uma vez que os processos subjacentes às contrações de Fe são considerados não estacionários, é necessário a modelação da componente $\mu(\mathbf{s})$ através de uma superfície de tendência linear para todas as campanhas. Para a campanha de 2002 a hipótese da isotropia é verificada; no entanto, para as outras duas é imprescindível a transformação das coordenadas através dos parâmetros $(\hat{\psi}_A, \hat{\psi}_R) = (0.00, 1.43)$ e $(\hat{\psi}_A, \hat{\psi}_R) = (0.00, 1.67)$, para as campanhas de 1992 e 1997, respetivamente.

Para a modelação do variograma, consideraram-se os modelos esférico com parâmetros $(\hat{\sigma}^2, \hat{\phi}, \hat{\tau}^2) = (914531.26, 92.61, 195346.01)$ para a campanha de 1992; e o modelo exponencial com parâmetros $(\hat{\sigma}^2, \hat{\phi}, \hat{\tau}^2) = (2551297.02, 44.78, 144600.77)$ e $(\hat{\sigma}^2, \hat{\phi}, \hat{\tau}^2) = (203747.01, 45.12, 129960.70)$ para as respetivas campanhas de 1997 e 2002. Aplicou-se, assim, o método *kriging* universal, obtendo-se os mapas de predição presentes na Figura 4.6.

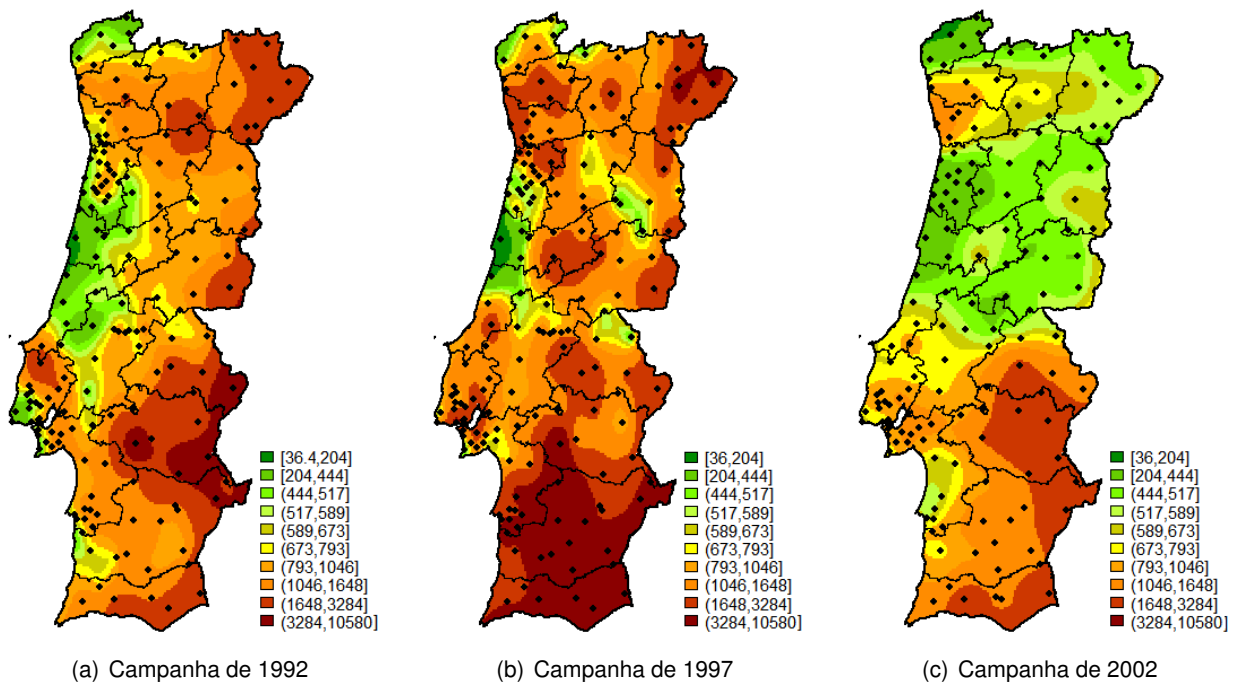


Figura 4.6: Superfícies preditas das concentrações de Fe nos musgos observadas em cada campanha.

O padrão de distribuição da deposição do Fe nos musgos sugere que os valores mais elevados estão contidos no sul do país, onde o coberto vegetal é menor e o clima mais seco, estando os solos mais expostos a fatores de erosão. Nas áreas correspondentes às localidades de Trás-os-Montes e Alentejo interior, a contaminação dos solos através das minas (tais como Alcáçovas, Moncorvo, etc.) é evidente. Verifica-se também que os baixos valores deste elemento apresentam-se em zonas calcárias, nomeadamente no Maciço Calcário Estremenho e no distrito de Leiria. Algumas fontes industriais podem ser causadoras de contaminação, nomeadamente nas regiões de Lisboa e Porto. Parece, ainda, existir uma diminuição destes valores na região Norte e Centro ao longo das três campanhas.

Manganês

Através da Tabela 15 (Anexo IV), verifica-se que para campanha de 2002 o processo subjacente é estacionário; no entanto, os processos relativos às campanhas de 1992 e 2006 são não estacionários, sendo desta forma, necessário a modelação da componente média através de uma superfície de tendência linear. Precisa-se, ainda, da verificação da isotropia do processo, o que acontece nas campanhas de 2002 e 2006. Para a campanha de 1992, uma vez que o processo é anisotrópico, é necessário a transformação das coordenadas de acordo com os parâmetros $(\hat{\psi}_A, \hat{\psi}_R) = (0.00, 1.24)$.

Assim, aplicou-se o método *kriging* universal às campanhas de 1992 e 2006 e *kriging* ordinário à campanha de 2002. Para tal, especificaram-se como funções de correlação dos variogramas das várias campanhas o modelo exponencial com parâmetros $(\hat{\sigma}^2, \hat{\phi}, \hat{\tau}^2) = (12814.36, 100.41, 6330.96)$ (campanha de 1992), $(\hat{\sigma}^2, \hat{\phi}, \hat{\tau}^2) = (6502.12, 32.25, 1875.27)$ (campanha de 2002) e $(\hat{\sigma}^2, \hat{\phi}, \hat{\tau}^2) = (3748.91, 7.84, 5236.40)$ (campanha de 2006).

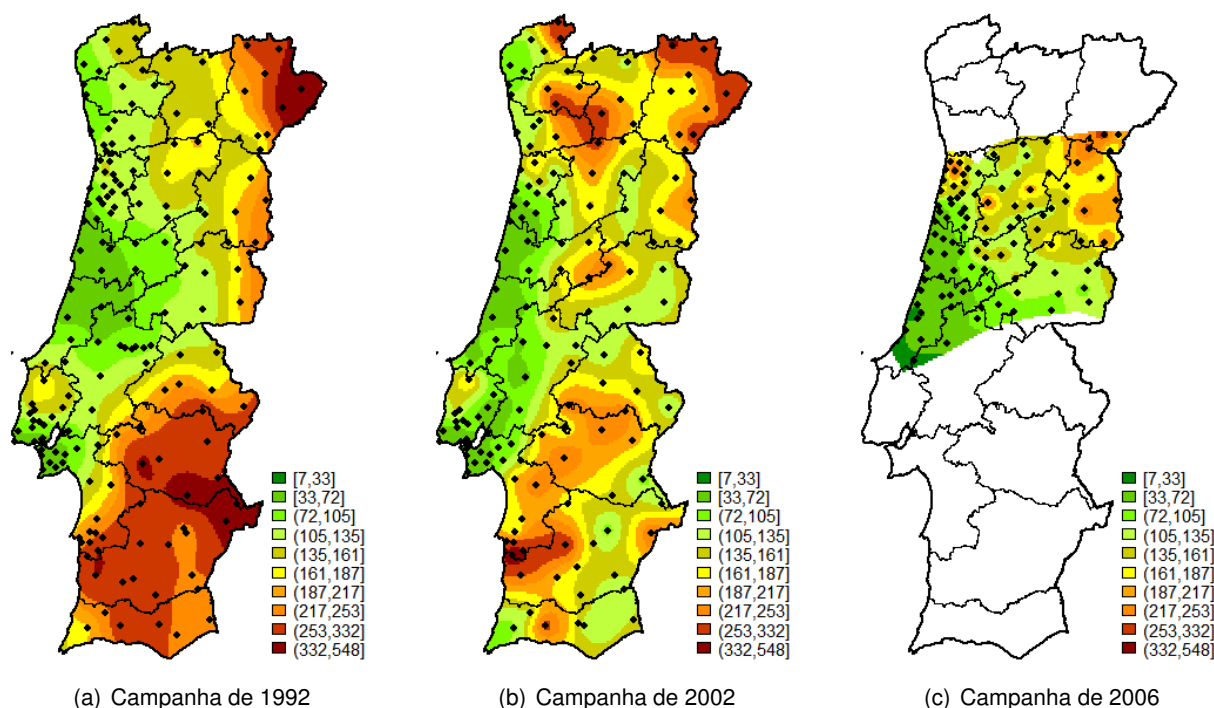


Figura 4.7: Superfícies preditas das concentrações de Mn nos musgos observadas em cada campanha.

Através da análise da Figura 4.7, verifica-se uma diminuição das concentrações de Mn ao longo das duas primeiras campanhas, nomeadamente no sul do país. No entanto, parece existir um alastramento das elevadas concentrações deste elemento nas regiões Norte e Centro.

A distribuição do Mn indica maior contaminação nas áreas mais secas e com menor coberto vegetal, à semelhança do Fe, nomeadamente na região transmontana e do Alentejo. Estes valores elevados nas zonas interiores devem corresponder, desta forma, à ação da erosão existente nos solos, além da forte componente geológica e presença de algumas minas (Mogadouro, Jales, Barrancos, Alcáçovas e Cercal).

Níquel

Para a aplicação da geoestatística, necessita-se da verificação da estacionaridade e da isotropia dos processos subjacentes aos dados. Para ambas as campanhas deste elemento tem-se a não estacionariedade dos processos. Portanto, modelou-se a componente média através de uma superfície de tendência linear. Relativamente à campanha de 2002 é verificada a hipótese de isotropia; no entanto, no caso da campanha de 1992 foi necessário a transformação das coordenadas através dos parâmetros $(\hat{\psi}_A, \hat{\psi}_R) = (0.00, 1.09)$. Selecionou-se o modelo exponencial, para as respetivas campanhas, com parâmetros $(\hat{\sigma}^2, \hat{\phi}, \hat{\tau}^2) = (2.34, 61.38, 0.75)$ e $(\hat{\sigma}^2, \hat{\phi}, \hat{\tau}^2) = (0.86, 38.94, 1.34)$, os quais foram fundamentais para a aplicação do método de *kriging* universal, obtendo-se os mapas de predição presentes na Figura 4.8.

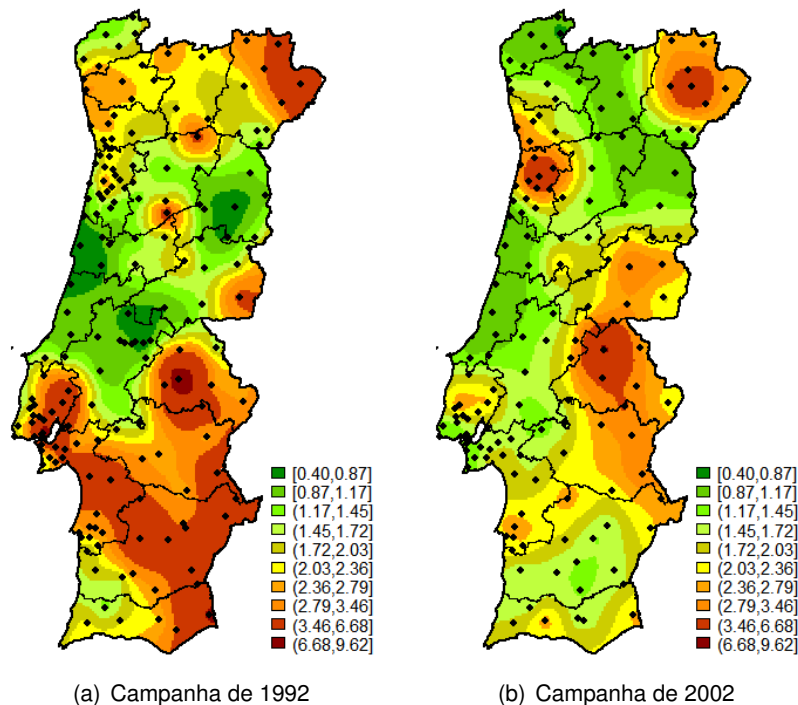


Figura 4.8: Superfícies preditas das concentrações de Ni nos musgos observadas em cada campanha.

Avaliando a Figura 4.8, verifica-se, para a campanha de 1992, a existência de uma mancha com valores elevados deste elemento na zona de Avis-Estremoz e na região da mina de Jales, bem como no estuário do Sado, onde os ventos dominantes podem ter importância no transporte de partículas das indústrias a norte.

No que diz respeito à campanha de 2002, verifica-se valores elevados na região do Porto, cujas contaminações podem resultar de emissões industriais, nomeadamente refinarias, indústrias químicas e centrais termo-elétricas da queima de carvão; na região norte de Lisboa, devido à presença de uma sucata de automóveis e fabricação de baterias; na região transmontana, onde os afloramentos de serpentinas originam a contaminação de Ni nesta região; e na região interior sul, junto à fronteira, possivelmente devido à atividade mineira.

Chumbo

Para ambas as campanhas, a componente da tendência espacial foi considerada estacionária. No que diz respeito à campanha de 2002, com o objetivo de que este processo seja isotrópico, considera-se a transformação das coordenadas através dos parâmetros $(\hat{\psi}_A, \hat{\psi}_R) = (0.00, 1.60)$. O processo relativo à campanha de 1997 é considerado isotrópico.

De forma a aplicarmos o método *kriging* ordinário, consideraram-se o modelo esférico com parâmetros $(\hat{\sigma}^2, \hat{\phi}, \hat{\tau}^2) = (139.17, 34.40, 89.97)$ para a campanha de 1992, e o modelo exponencial com parâmetros $(\hat{\sigma}^2, \hat{\phi}, \hat{\tau}^2) = (90.19, 41.38, 20.80)$ para a campanha de 2002, como sendo as funções de correlação dos variogramas teóricos.

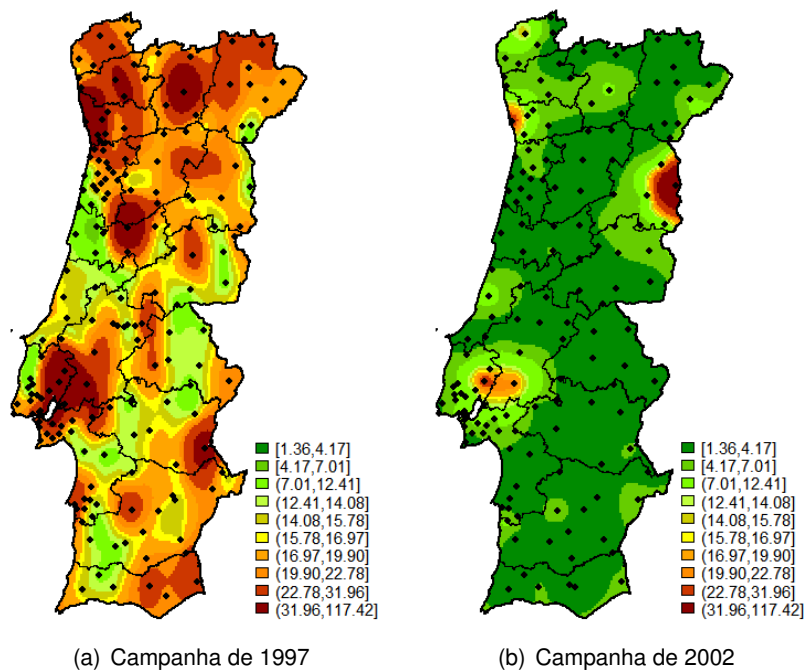


Figura 4.9: Superfícies previstas das concentrações de Pb nos musgos observadas em cada campanha.

Na campanha de 1997 verifica-se a presença de valores elevados nas áreas mais densamente povoadas do país, cujo tráfego rodoviário é mais intenso; na região de Lisboa estendendo-se para norte através do vale do Tejo, como resultado da influência de indústrias metalúrgicas ou em resultado do uso de inseticidas na agricultura; alguns pontos no Norte, que correspondem a novos nós rodoviários.

No que diz respeito à campanha de 2002, o padrão espacial deste elemento indica que os valores mais elevados estão situados nas áreas urbanas e industriais, nomeadamente em Lisboa e Porto. Analisando a Figura 4.9, evidencia-se um notório decréscimo dos valores de Pb da campanha de 1997 para a de 2002, facto que pode ser explicado pela substituição total da gasolina com chumbo por gasolina sem chumbo, que ocorrera no ano de 1999. O valor mais elevado para esta campanha foi encontrado próximo do IP5 (Vilar Formoso), junto à fronteira. Nas áreas rurais, uma das fontes de contaminação poderá resultar dos tiros de caça ou do abandono dos cartuchos de caça.

Zinco

O processo subjacente às concentrações de Zn nos musgos da campanha de 1992 é não estacionário, modelando-se a componente da média por um modelo de superfície de tendência linear; sendo considerado um processo isotrópico. Verifica-se, através da Tabela 24 (Anexo IV), que o processo relativo à campanha de 2002 é considerado estacionário mas anisotrópico, procedendo-se, desta forma, a uma transformação das coordenadas através dos parâmetros $(\hat{\psi}_A, \hat{\psi}_R) = (0.14, 2.43)$.

Posteriormente, os modelos esférico e exponencial foram os selecionados para modelar as concentrações de Zn nas campanhas de 1992 e 2002, respetivamente. As estimativas para os parâmetros dos variogramas das campanhas de 1992 e 2002 são $(\hat{\sigma}^2, \hat{\phi}, \hat{\tau}^2) = (228.97, 40.99, 156.86)$ e $(\hat{\sigma}^2, \hat{\phi}, \hat{\tau}^2) = (120.11, 38.50, 125.00)$, respetivamente. Desta forma, procedeu-se à aplicação do método *kriging* universal para a campanha de 1992 e do método de *kriging* ordinário para a campanha de 2002, obtendo-se os mapas de predição apresentados na Figura 4.10.

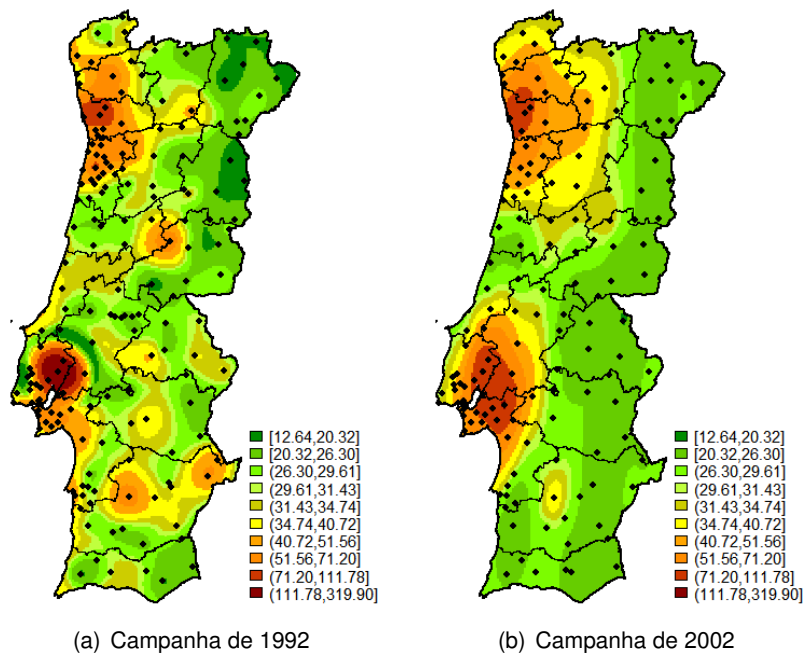


Figura 4.10: Superfícies preditas das concentrações de Zn nos musgos observadas em cada campanha.

Verifica-se, através da Figura 4.10, que as concentrações mais elevadas deste elemento químico apresentam-se nos distritos mais populosos e industrializados tais como Lisboa/Sines (onde se encontram instaladas várias indústrias) e Porto/Aveiro (existindo refinarias e indústrias metalúrgicas), em ambas as campanhas.

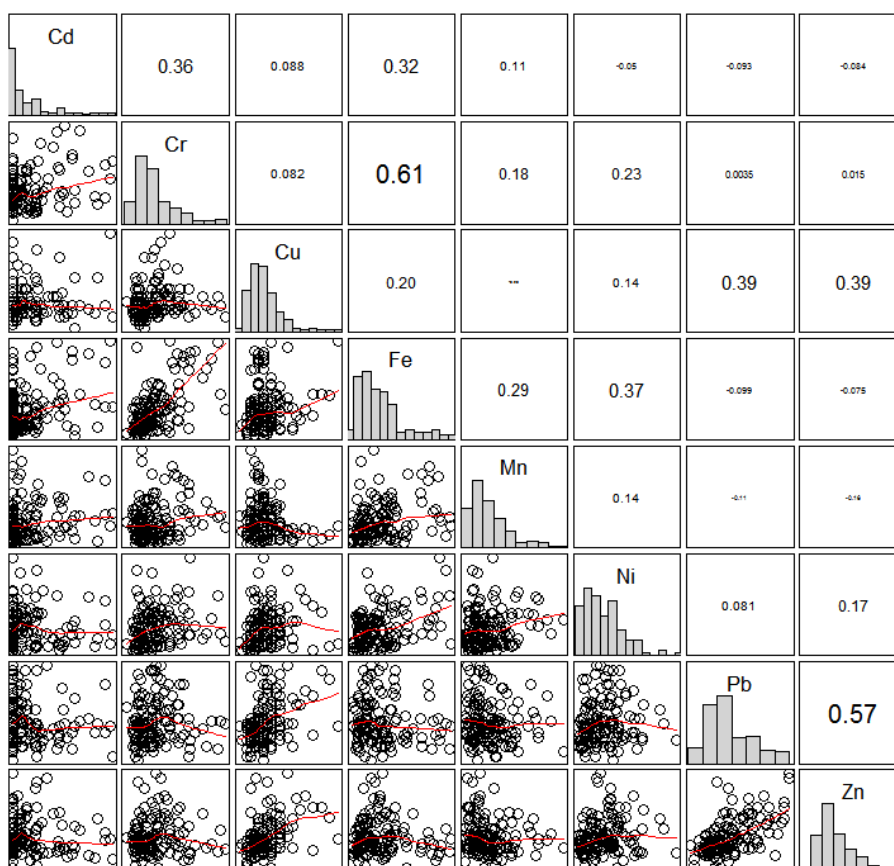
Observa-se, para a campanha de 1992, a presença de valores elevados nas minas situadas no interior do país (tais como, as minas de Panasqueira e Aljustrel). Existem outros pontos que apresentam valores elevados no interior do país que podem corresponder à ação da erosão. Em geral, nas zonas urbanas é necessário considerar o tráfego automóvel como fonte de contaminação ambiental deste elemento.

4.2 Reconhecimento dos locais poluídos e das relações subjacentes aos elementos químicos - análise multivariada

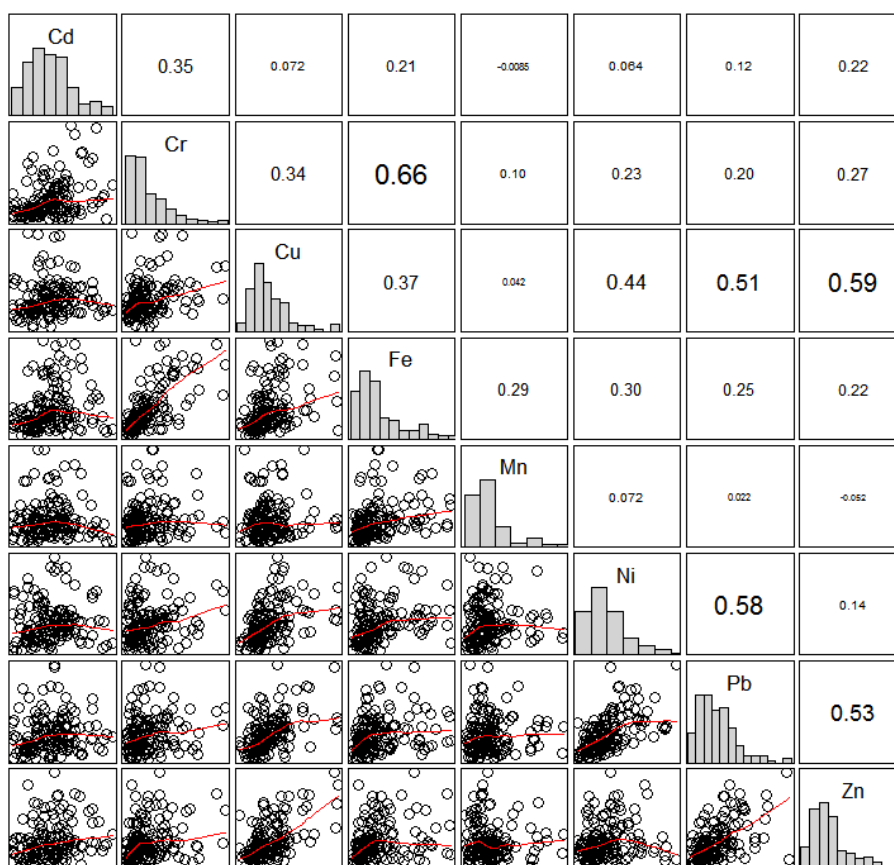
Um estudo de biomonitorização é tanto mais completo quanto melhor procurar compreender e modelar as interações dos vários elementos químicos. Esta atitude implica a aplicação de metodologias de análise multivariada capazes de identificar, descrever e modelar essas relações. Desta forma, são aplicadas diversas metodologias de análise multivariada como análise em componentes principais, análise de *clusters* e análise fatorial (presente no capítulo 5) descritas, teoricamente, no capítulo 3.

Um dos primeiros passos na análise das relações existentes entre os diversos elementos químicos consiste no cálculo das correlações entre cada par de elementos. Neste sentido, construíram-se diagramas de dispersão entre os vários elementos químicos, para cada campanha, aferindo-se os valores das respetivas correlações de *Pearson*. O cálculo das correlações entre os elementos foi realizado eliminando os *outliers* severos, uma vez que estes valores distorciam o resultado final.

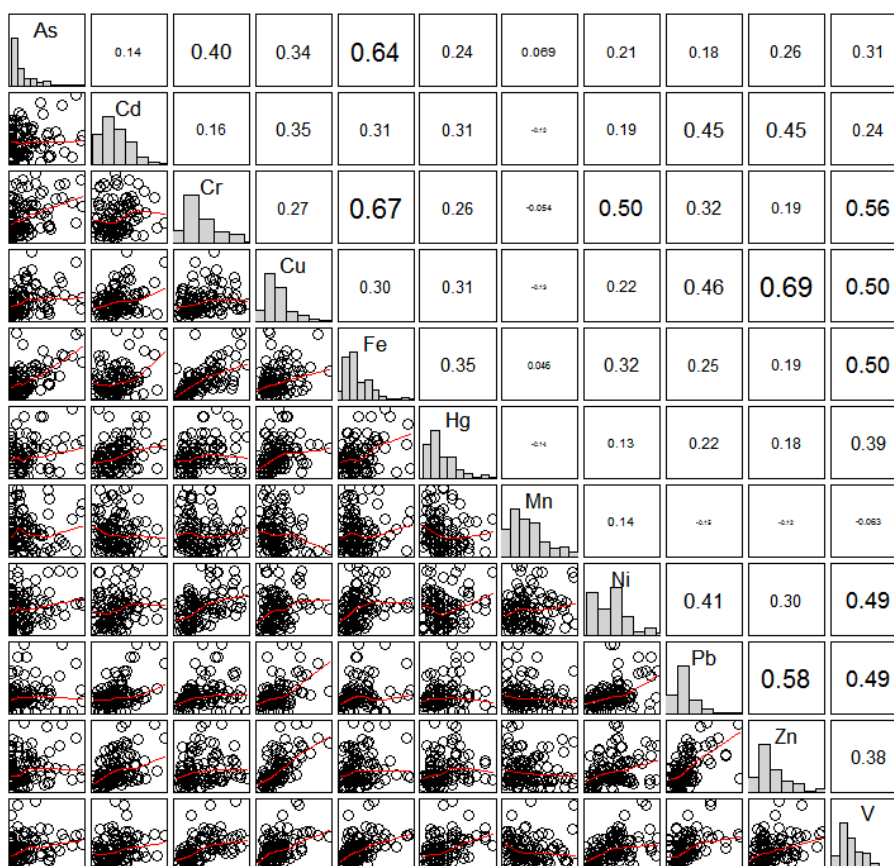
Os vários diagramas estão apresentados nas Figuras 4.11(a), 4.11(b), 4.11(c) e 4.11(d). Nestas Figuras, os valores mais elevados, no que diz respeito ao valor da correlação de *Pearson*, encontram-se a um tamanho de letra maior, sendo estatisticamente significativas as correlações entre estes pares uma vez que apresentam um $p - \text{valor} < 0.05$. Nestas, é ainda possível observarmos os histogramas de cada metal.



(a) Campanha de 1992



(b) Campanha de 1997



(c) Campanha de 2006

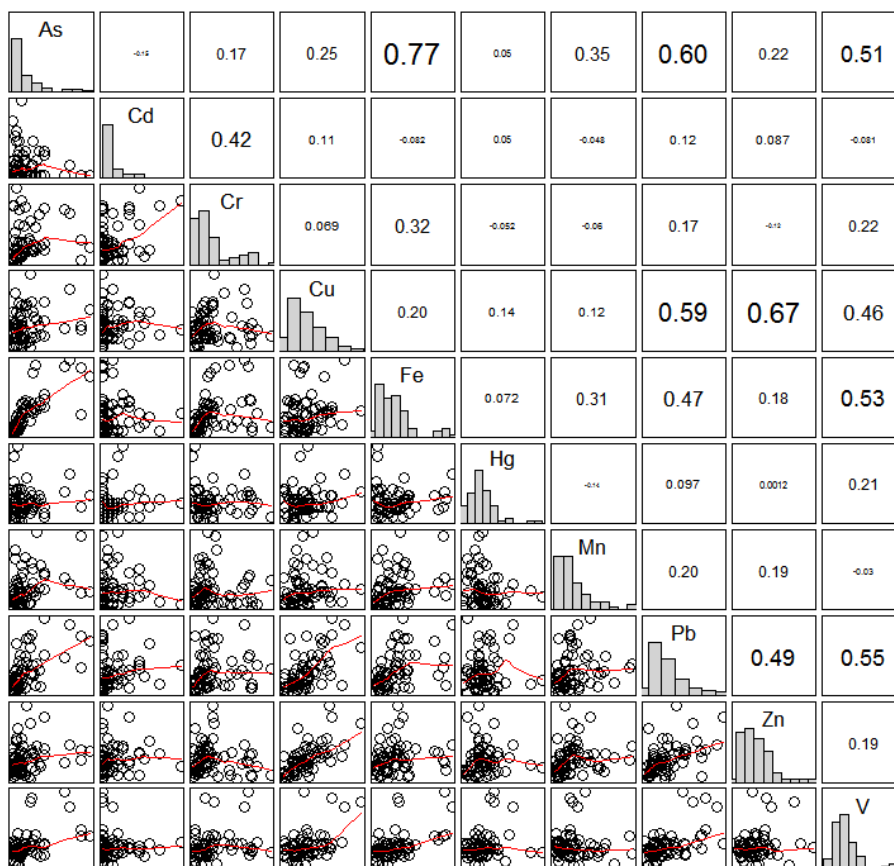


Figura 4.11: Matrices das correlações entre as concentrações dos vários elementos químicos nos musgos observadas em cada campanha. Representam-se os histogramas (diagonal) dos elementos químicos, os coeficientes de correlação (acima da diagonal) e os diagramas de dispersão (abaixo da diagonal).

Verifica-se, através da observação das Figuras 4.11(a), 4.11(b), 4.11(c) e 4.11(d), que ao longo das campanhas existe uma forte associação entre os elementos Cd/Cr, Cr/Fe, Zn/Cu, Pb/Cu, Fe/As e ainda entre o Pb/Zn.

Existem valores de correlações que apesar de não serem muito elevados, podem ser um indicador da existência de fontes de contaminação comuns para alguns dos pares de elementos químicos.

4.2.1 Análise exploratória dos locais mais poluídos em Portugal - análise de *clusters*

O objetivo final da aplicação da AC sobre as matrizes de dados das várias campanhas é identificar, na região de Portugal continental, localizações com características semelhantes ao nível das concentrações dos elementos químicos nos musgos. Consequentemente, ter-se-á a informação de quais destas localizações são as mais poluídas.

A forma mais usual para representar o resultado final dos diversos grupos ou *clusters* é o chamado dendrograma, representados na Figura 4.12. Esta Figura contém os dendrogramas obtidos para as quatro campanhas em estudo.

Como estes gráficos são uma simplificação em duas dimensões de relações n -dimensionais, é inevitável que algumas distorções apareçam quanto à (dis)semelhança. A medida de tal distorção pode ser obtida por um coeficiente de correlação, designado por correlação *cofonética*, entre os valores da

matriz de dissimilaridades inicial e a matriz de dissimilaridades dos derivados do dendrograma. A escolha das diferentes distâncias e métodos devem ser alvo de bom senso, uma vez que o fenómeno em causa apresenta uma grande variabilidade. Por esta razão, a escolha de determinados métodos, apesar do valor da correlação *cofonética* ser elevado, pode implicar resultados completamente distorcidos da realidade. Para a análise das várias campanhas, utiliza-se a medida de dissimilaridade máxima, usualmente designada por dissimilaridade de *Chebychev*, e o método de ligação *Ward*. Os valores das correlações *cofonéticas*, para as várias campanhas, estão apresentados na Tabela 4.5.

Tabela 4.5: Valores das correlações *cofonéticas*, para as várias campanhas.

Campanha			
1992	1997	2002	2006
Coeficiente	Coeficiente	Coeficiente	Coeficiente
0.68	0.76	0.74	0.83

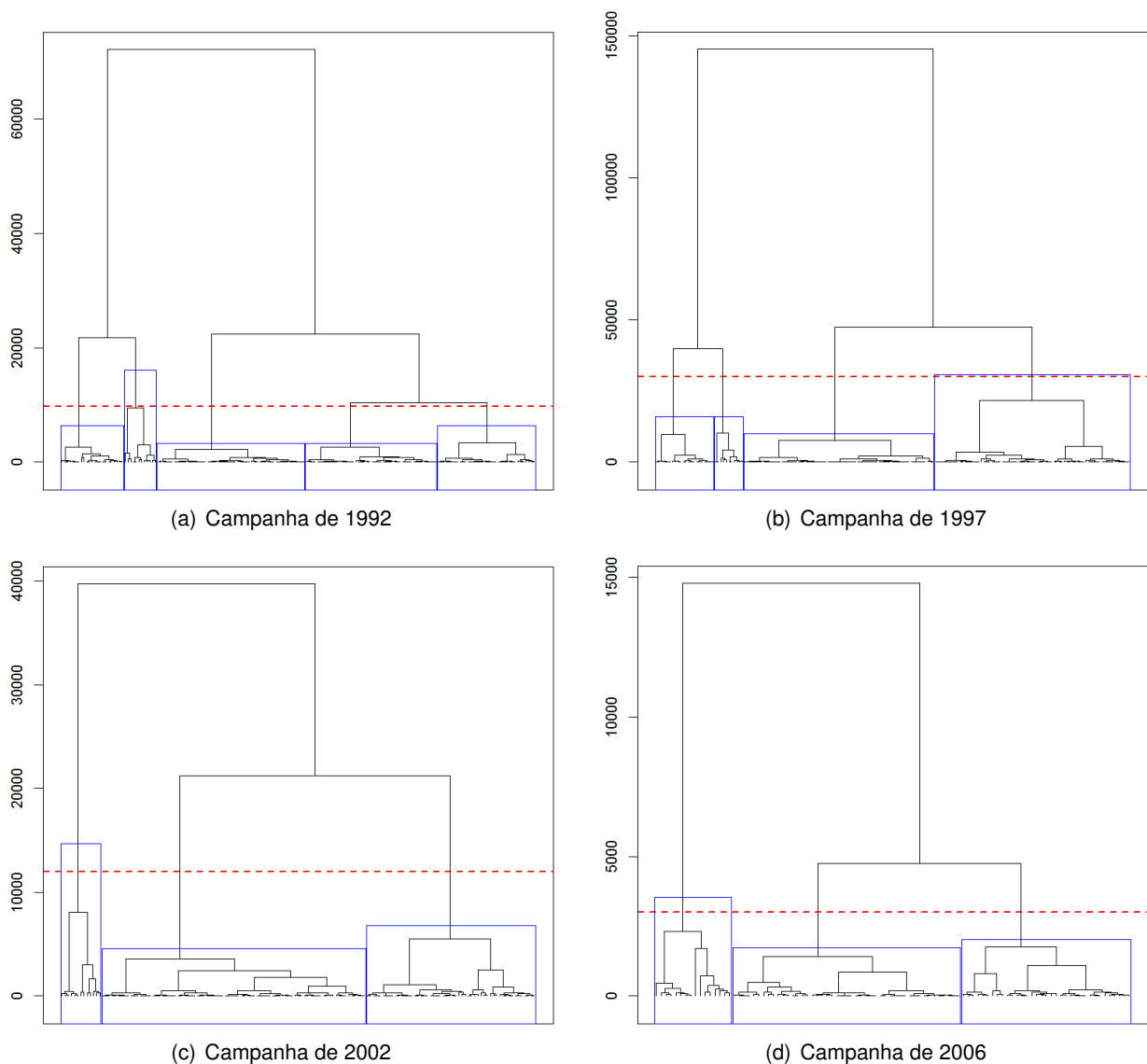


Figura 4.12: Dendrogramas das concentrações dos vários elementos químicos nos musgos observadas em cada campanha, utilizando o método de agrupamento de *Ward* e a medida de dissimilaridade de *Chebychev*.

A questão que se coloca neste momento é quantos grupos reter, de modo a que este número seja o ideal. Nesta decisão reter poucos grupos pode levar a que estes sejam demasiado heterogéneos e demasiados grupos tornam a sua interpretação difícil. Desta forma, optou-se pela utilização do método empírico por ser o mais frequente, permitindo que esta divisão se aproxime mais da realidade. Através da análise da Figura 4.12, verifica-se a presença de cinco grupos na campanha de 1992, quatro grupos na campanha de 1997 e três grupos nas campanhas de 2002 e 2006.

Consequentemente, obteve-se a divisão da região de Portugal continental pelos diferentes grupos, nas várias campanhas, assinalados na Figura 4.13.

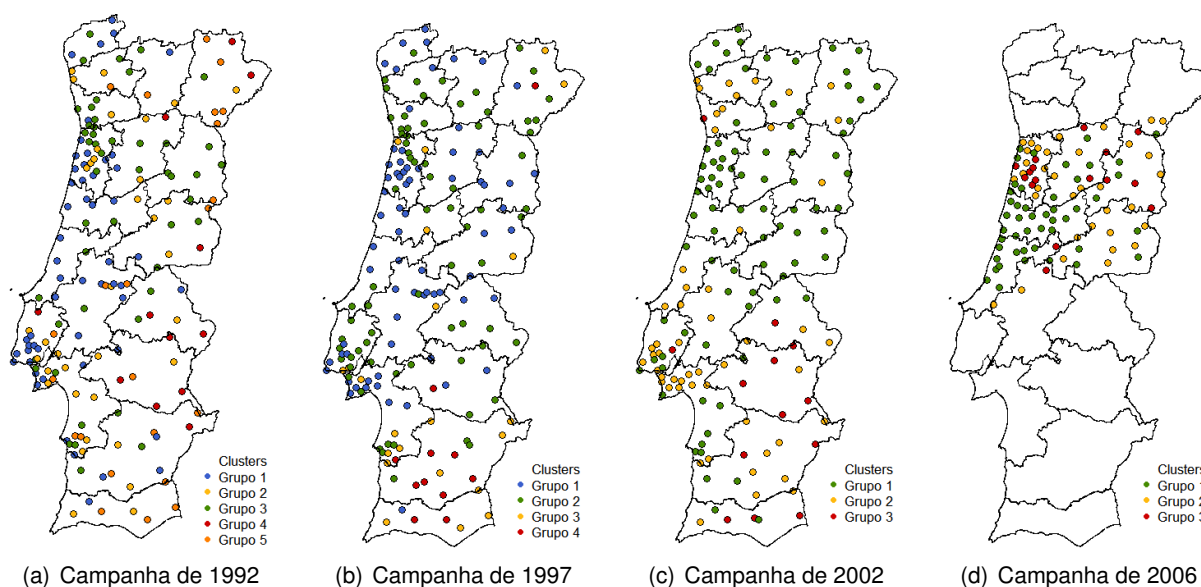


Figura 4.13: Mapas de Portugal continental com as diferentes localizações amostradas em cada campanha subdivididas pelos vários grupos.

Conhecido o número de grupos presentes em cada campanha, pretende-se compreender se existem diferenças significativas entre eles, tendo em consideração que as variáveis traduzem a qualidade do ambiente. Para testar a existência dessas diferenças recorreu-se ao teste não paramétrico de Kruskal-Wallis³, face à privação da normalidade dos grupos.

Os resultados da aplicação deste teste, assumindo-se como hipótese nula que não existem diferenças significativas entre os vários grupos relativamente às concentrações dos elementos químicos nos musgos, encontram-se na Tabela 4.6. Nesta Tabela estão, ainda, registados os valores médios das concentrações dos elementos químicos nos musgos, as quais foram subdivididas pelos respetivos grupos.

³O teste de *Kruskal-Wallis* é um teste não paramétrico utilizado para comparar três ou mais amostras, cuja hipótese nula é de que as k populações possuem funções de distribuição iguais contra a hipótese alternativa de que pelo menos duas das populações possuem funções de distribuição diferentes, sendo a estatística de teste dada por

$$ET = \frac{12}{n(n+1)} \sum_{i=1}^k \frac{r_i^2}{n_i} - 3(n+1),$$

onde cada amostra contém n_i observações, $n = \sum_{i=1}^k n_i$ e r_i é a soma dos números de ordem das n_i observações da amostra i .

Relativamente à campanha de 1992, pode-se verificar que:

- consultando os valores de prova obtidos através do teste referido, conclui-se que existe evidência estatística suficiente nos dados para rejeitar a igualdade dos valores médios nos cinco grupos em todos os elementos químicos, à exceção do Pb e do Zn;
- o grupo 1 é aquele em que todos os elementos químicos apresentam um valor médio mais baixo. Portanto, pode-se concluir que este é o grupo das localizações que se encontram menos poluídas;
- o grupo 2 destaca-se, essencialmente, pelo elevado valor médio do elemento químico Cu, estando associado a origens antropogénicas, nomeadamente a indústrias de processamento de metais e ao tráfego rodoviário. Os elementos deste grupo apresentam um valor médio mais elevado que os grupos 1 e 3;
- o grupo 3 não se destaca por nenhum elemento químico. Os elementos químicos deste grupo apresentam um valor médio mais elevado que o grupo 1. Desta forma, este grupo pode ser considerado pouco poluído;
- os elementos Cd, Cr, Fe, Mn e Ni, presentes no grupo 4, são os que apresentam um valor médio mais elevado. Desta forma, este grupo é representado por localizações mais secas e com menor cobertura vegetal, características do Baixo Alentejo;
- os elementos químicos, do grupo 5, apresentam um valor médio mais baixo que o grupo 4. Através da Figura 4.13(a), verifica-se que enquanto que todas as localizações do grupo 4 concentram-se no Baixo Alentejo; as localizações deste grupo, para além da região referida, encontram-se também localizados noutras zonas rurais do país, como Trás-os-Montes e Algarve.

No que concerne à campanha de 1997, faz-se notar que:

- consultando os valores de prova obtidos através do teste referido, conclui-se que existe evidência estatística suficiente nos dados para rejeitar a igualdade dos valores médios nos quatro grupos em todos os elementos químicos;
- os elementos químicos do grupo 1 apresentam um valor médio mais baixo. Logo, pode-se concluir que este grupo é constituído pelas localizações menos poluídas;
- o grupo 2 apresenta um valor médio elevado para os elementos químicos Cd, Pb e Zn, tendo estes elementos fontes, essencialmente, antropogénicas. Desta forma, as localizações que pertencem a este grupo estão presentes, maioritariamente, nas áreas metropolitanas de Lisboa e Porto;
- os elementos químicos do grupo 3 apresentam um valor médio mais elevado que os grupos 1 e 2, com exceção dos elementos Cd, Ni, Pb e Zn;

- o grupo 4 é aquele em que quase todos os elementos químicos apresentam um valor médio mais elevado. No entanto, neste grupo os elementos Pb e Zn apresentam um valor médio mais baixo.

No que diz respeito à campanha de 2002, verifica-se que:

- consultando os valores de prova obtidos através do teste referido, conclui-se que existe evidência estatística suficiente nos dados para rejeitar a igualdade dos valores médios nos três grupos em todos os elementos químicos, à exceção dos elementos Cd e Zn;
- os elementos químicos do grupo 1 apresentam um valor médio mais baixo, sendo por isso o grupo que é constituído pelas localizações que apresentam menos poluição;
- os elementos químicos do grupo 2 apresentam um valor médio mais elevado que o grupo 1 e mais baixo do que o grupo 3. No entanto, o elemento Hg apresenta um valor médio mais elevado para este grupo, sendo que este elemento químico está associado a diversos ramos de atividade humana como mineração (amalgamento do ouro), indústrias cerâmicas e farmacêuticas, instrumentos de medição, etc.;
- o grupo 3 é aquele em que os elementos químicos apresentam um valor médio mais elevado, à exceção do Hg. Assim, pode-se considerar que este grupo contém as localizações mais poluídas. Note-se que este grupo apresenta um valor médio muito elevado para os elementos Fe e Mn, sendo esta a razão das localizações deste grupo se situarem no Alentejo e Algarve. É de notar, ainda, que o valor médio do Pb desceu muito em relação às campanhas de 1992 e 1997, consequência da substituição total da gasolina com chumbo por gasolina sem chumbo.

Relativamente à campanha de 2006, verifica-se que:

- consultando os valores de prova obtidos através do teste referido, conclui-se que existe evidência estatística suficiente nos dados para rejeitar a igualdade dos valores médios nos três grupos em todos os elementos químicos, à exceção dos elementos Cd e Hg;
- os elementos químicos do grupo 1 apresentam um valor médio mais baixo, sobressaindo-se o valor médio do Cr, o qual encontra-se tipicamente associado ao transporte erosivo do solo por ação do vento;
- o grupo 2 é aquele em que os elementos químicos apresentam um valor médio mais elevado que o grupo 1, à exceção do elemento químico Cr;
- os elementos químicos do grupo 3 apresentam um valor médio mais elevado, sendo por isso o grupo que contém as localizações mais poluídas. Este grupo sobressai-se, essencialmente, pelo elevado valor médio dos elementos As, Fe e V.

Tabela 4.6: Valores médio dos diferentes grupos subjacentes a cada campanha e respetivos valores de prova do teste de Kruskal-Wallis.

Campanha		Grupos					<i>p – valor</i>
		Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5	
1992	Cd	0.10	0.21	0.21	0.58	0.32	0.026
	Cr	1.05	1.92	1.37	8.75	3.83	< 0.001
	Cu	6.97	12.78	8.15	10.92	11.17	0.002
	Fe	370.43	1204.89	748.10	4440.17	2116.68	< 0.001
	Mn	97.83	158.56	140.56	373.50	225.30	< 0.001
	Ni	1.40	3.31	1.77	5.84	3.36	< 0.001
	Pb	18.56	17.83	15.52	13.63	20.35	0.414
	Zn	32.87	41.45	40.42	30.58	54.61	0.065
1997	Cd	0.71	0.85	0.76	0.81		0.011
	Cr	1.56	3.07	6.03	19.65		< 0.001
	Cu	6.35	9.70	12.46	13.13		< 0.001
	Fe	571.12	1696.22	4077.38	8120.59		< 0.001
	Mn	120.25	211.45	235.98	378.02		< 0.001
	Ni	8.48	15.24	14.86	32.36		< 0.001
	Pb	14.91	28.84	22.84	16.62		< 0.001
	Zn	39.21	66.42	49.99	49.89		0.002
2002	As	0.35	0.98	2.75			< 0.001
	Cd	0.45	0.52	0.75			0.096
	Cr	1.08	1.67	5.41			< 0.001
	Cu	5.82	8.26	13.32			< 0.001
	Fe	402.99	963.54	3073.37			< 0.001
	Hg	0.06	0.13	0.07			0.005
	Mn	148.86	128.86	215.08			0.012
	Ni	1.94	1.71	3.32			< 0.001
	Pb	4.98	5.58	7.34			0.012
	Zn	30.49	48.95	44.95			0.066
	V	2.43	4.46	7.78			< 0.001
2006	As	0.32	1.52	3.31			< 0.001
	Cd	0.03	0.03	0.09			0.369
	Cr	0.66	0.64	1.91			< 0.001
	Cu	3.39	4.74	7.06			0.001
	Fe	186.22	412.54	1092.83			< 0.001
	Hg	0.07	0.06	0.12			0.051
	Mn	83.18	140.79	174.04			0.002
	Pb	1.54	2.12	2.65			< 0.001
	Zn	23.81	31.32	49.09			0.002
	V	5.72	7.53	72.80			< 0.001

4.2.2 Análise exploratória das relações existentes entre os elementos químicos - análise em componentes principais

Com o intuito de sintetizar e analisar a informação disponível, nomeadamente das relações existentes entre os vários elementos químicos, utiliza-se a metodologia de análise multivariada ACP. O objetivo mais imediato desta metodologia é verificar se existe um pequeno número das primeiras componentes principais que seja responsável por explicar uma proporção elevada da variabilidade total associada ao conjunto original de dados.

A adequação desta técnica aos dados em questão é avaliada através do teste de esfericidade de Bartlett⁴, cuja hipótese nula é: a matriz de correlação é uma matriz identidade, indicando a ausência de correlação entre as várias variáveis em estudo.

Consultando a Tabela 4.7, verifica-se que os elementos químicos encontram-se correlacionados ($p - valor < 0.05$), podendo-se também verificar este facto através dos valores dos determinantes das matrizes de correlação que são próximos de zero. Assim, pode-se aplicar a ACP para obter posteriormente novas variáveis não correlacionadas.

Tabela 4.7: Valores dos determinantes (det) das matrizes de correlação subjacentes a cada campanha e respetivos valores de prova do teste de esfericidade de Bartlett.

Campanha							
1992		1997		2002		2006	
det	$p - valor$	det	$p - valor$	det	$p - valor$	det	$p - valor$
0.06	< 0.001	0.15	< 0.001	0.08	< 0.001	0.01	< 0.001

De seguida, procedeu-se à derivação das componentes principais sobre as matrizes dos dados, após a estandardização das mesmas, começando-se por calcular a proporção de variância explicada pelas novas variáveis e também o valor das correlações entre os elementos em estudo e as componentes principais, para cada campanha.

Nas Tabelas 4.8, 4.9, 4.10 e 4.11, relativas às várias campanhas, os valores singulares são relativos à matriz de correlação, coincidindo com a raiz quadrada dos valores próprios desta mesma matriz. Os *loadings* deste método correspondem aos coeficientes das combinações lineares de cada uma das componentes principais em função das variáveis originais, coincidindo com os vetores singulares ou vetores próprios da matriz de correlação.

Nesta metodologia, pode-se reter as componentes principais que permitam explicar entre 70% e

⁴O teste de esfericidade de Bartlett testa a hipótese de que as variáveis não sejam correlacionadas na população. A hipótese nula diz que a matriz de correlação da população é uma matriz identidade, o que indica que é inapropriado a aplicação da metodologia ACP e AF. A estatística de teste é dada por

$$ET = - \left[(n - 1) - \frac{2p + 5}{6} \right] \ln |\mathbf{R}|,$$

que tem uma distribuição χ^2 com $\nu = \frac{p(p-1)}{2}$ graus de liberdade, onde n é o número da amostra, p é o número de variáveis e $|\mathbf{R}|$ é o determinante da matriz de correlação.

90% da variabilidade total dos dados - critério da variância total explicada, ou cujos valores próprios sejam superiores a um (há quem considere acima de 0.7) - critério de *Kaiser*.

Através da análise da Tabela 4.8 (respetiva à campanha de 1992), pode-se reter as quatro primeiras componentes principais, as quais explicam aproximadamente 80 % da variabilidade total dos dados. Nesta Tabela também estão contidas as correlações entre os elementos químicos em estudo e as várias componentes principais, verificando-se que:

- todas as correlações da primeira componente principal (CP1) com os vários elementos químicos são positivas, indicando que esta componente principal não tem capacidade de discriminação entre os elementos. Verifica-se, ainda, um elevado valor das correlações entre esta componente e os elementos Cr e Fe (relação presente na análise bivariada dos elementos químicos da campanha de 1992, Figura 4.11(a));
- existe uma correlação muito forte da segunda componente principal (CP2) com os elementos químicos Pb e Zn. Note-se que existe uma elevada correlação entre os elementos Pb e Zn (ver Figura 4.11(a)), pelo que é natural que estes elementos se encontrem na mesma componente principal;
- os elementos químicos Cd e Mn são aqueles que apresentam uma maior correlação com a terceira componente principal (CP3), apresentando sinais contrários;
- no que diz respeito à quarta componente principal (CP4), é o elemento químico Cu que interfere mais na obtenção desta componente.

Tabela 4.8: *Loadings*, valores das correlações entre os elementos químicos e as componentes principais e respetivos valores da variância explicada por cada componente principal na campanha de 1992.

	Loadings				Correlações			
	CP1	CP2	CP3	CP4	CP1	CP2	CP3	CP4
Cd	0.33	-0.16	0.63	-0.33	0.53	-0.23	0.61	-0.29
Cr	0.52	-0.19	0.15	-0.08	0.83	-0.27	0.15	-0.07
Cu	0.22	0.32	0.43	0.74	0.35	0.46	0.42	0.65
Fe	0.50	-0.12	-0.23	0.26	0.80	-0.17	-0.22	0.23
Mn	0.36	-0.18	-0.57	0.18	0.57	-0.26	-0.56	0.16
Ni	0.40	0.15	-0.09	-0.43	0.64	0.21	-0.09	-0.38
Pb	0.09	0.64	-0.07	-0.10	0.15	0.91	-0.07	-0.09
Zn	0.14	0.61	-0.13	-0.19	0.22	0.87	-0.13	-0.16
Valores singulares	1.59	1.44	0.98	0.88				
Proporção de variância (%)	0.32	0.26	0.12	0.10				
Proporção acumulada (%)	0.32	0.58	0.70	0.79				

No que diz respeito à campanha de 1997, pode-se reter as quatro primeiras componentes principais, explicando-se 76% da variabilidade total dos dados, cujo valor próprio associado a esta quarta

componente é igual a um (ver Tabela 4.9). Relativamente aos valores das correlações entre os vários elementos químicos e as componentes principais apresentados na Tabela 4.9, conclui-se que:

- as correlações da primeira componente principal (CP1) com os vários elementos químicos são negativas. Assim, esta componente não apresenta uma capacidade de discriminação entre os elementos, destacando-se os elementos químicos Cr, Fe e Ni;
- relativamente à segunda componente principal (CP2), os valores das correlações desta componente com os elementos químicos Pb e Zn são elevados, relação patente na campanha de 1992;
- os elementos químicos que mais interferem na obtenção da terceira (CP3) e da quarta componente principal (CP4) são os elementos Mn e Cd, respetivamente.

Tabela 4.9: *Loadings*, valores das correlações entre os elementos químicos e as componentes principais e respetivos valores da variância explicada por cada componente principal na campanha de 1997.

	Loadings				Correlações			
	CP1	CP2	CP3	CP4	CP1	CP2	CP3	CP4
Cd	-0.06	-0.09	0.20	-0.95	-0.09	-0.12	0.22	-0.95
Cr	-0.46	0.40	0.31	0.04	-0.69	0.52	0.34	0.04
Cu	-0.34	-0.29	-0.26	0.13	-0.51	-0.37	-0.28	0.13
Fe	-0.42	0.20	-0.45	-0.13	-0.63	0.26	-0.48	-0.13
Mn	-0.27	0.14	-0.60	-0.16	-0.41	0.18	-0.65	-0.16
Ni	-0.45	0.33	0.42	0.14	-0.67	0.42	0.45	0.14
Pb	-0.35	-0.51	0.22	-0.00	-0.53	-0.66	0.24	-0.00
Zn	-0.30	-0.57	0.05	0.11	-0.46	-0.73	0.06	0.11
Valores singulares	1.51	1.29	1.08	1.00				
Proporção de variância (%)	0.28	0.21	0.14	0.12				
Proporção acumulada (%)	0.28	0.49	0.64	0.76				

Relativamente à campanha de 2002, reteve-se apenas as cinco primeiras componentes principais, explicando-se 71% da variabilidade total dos dados (ver Tabela 4.10). Note-se que o valor próprio associado à quinta componente principal é aproximadamente igual a um, como se pode observar pela Tabela 4.10. Analisando os valores das correlações entre os vários elementos químicos e as componentes principais, presentes na Tabela 4.10, evidencia-se que:

- mais uma vez, a primeira componente principal (CP1) não tem um carácter discriminativo. Os elementos químicos Cr, Cu e Fe são os que mais influenciam a obtenção desta componente principal;
- verifica-se valores elevados para as correlações entre a segunda componente principal (CP2) e os elementos químicos Mn e Zn, exibindo sinais contrários;

- no que diz respeito à terceira componente principal (CP3), os elementos químicos que apresentam maior peso nesta componente são o Ni, As e Mn;
- os elementos químicos Hg e Cd são os que mais influenciam a quarta (CP4) e a quinta componente principal (CP5), respetivamente.

Tabela 4.10: *Loadings*, valores das correlações entre os elementos químicos e as componentes principais e respetivos valores da variância explicada por cada componente principal na campanha de 2002.

	<i>Loadings</i>					<i>Correlações</i>				
	CP1	CP2	CP3	CP4	CP5	CP1	CP2	CP3	CP4	CP5
As	0.29	-0.36	0.42	-0.17	0.11	0.50	-0.45	0.46	-0.18	0.11
Cd	0.12	0.24	0.21	-0.52	-0.57	0.20	0.30	0.23	-0.54	-0.56
Cr	0.41	-0.24	-0.32	-0.09	0.16	0.70	-0.30	-0.34	-0.09	0.16
Cu	0.39	0.28	0.08	0.23	0.14	0.67	0.35	0.08	0.24	0.14
Fe	0.43	-0.33	0.12	-0.07	0.20	0.74	-0.41	0.13	-0.08	0.20
Hg	0.04	-0.03	0.05	0.76	-0.37	0.06	-0.04	0.06	0.81	-0.37
Mn	0.04	-0.44	0.45	0.14	-0.40	0.07	-0.55	0.48	0.15	-0.40
Ni	0.26	-0.09	-0.52	-0.10	-0.51	0.44	-0.11	-0.57	-0.10	-0.51
Pb	0.19	0.35	0.28	0.10	0.12	0.33	0.45	0.30	0.11	0.12
Zn	0.28	0.49	0.23	-0.03	-0.07	0.48	0.61	0.25	-0.04	-0.07
V	0.47	0.09	-0.22	0.15	-0.03	0.81	0.11	-0.23	0.15	-0.03
Valores singulares	1.72	1.26	1.08	1.06	0.99					
Proporção de variância (%)	0.27	0.14	0.11	0.10	0.09					
Proporção acumulada (%)	0.27	0.41	0.52	0.62	0.71					

Através da análise da Tabela 4.11 (relativa à campanha de 2006), pode-se reter as quatro primeiras componentes principais, explicando-se 75% da variabilidade total dos dados, cujo valor próprio associado à quarta componente principal é, aproximadamente, igual a um. Observando as correlações entre os vários elementos químicos e as componentes principais apresentadas na Tabela 4.11, verifica-se que:

- à semelhança das restantes campanhas, a primeira componente principal (CP1) não tem capacidade de discriminação entre os elementos químicos, uma vez que todas as correlações entre as componentes principais e os elementos são positivas. Os elementos Cd, Cr e Cu são os que têm mais influência nesta componente;
- relativamente à segunda componente principal (CP2), os elementos químicos que apresentam maior correlação com esta componente são os elementos As, Fe e Mn;
- os elementos químicos Cd, Hg e V são os que mais influenciam a obtenção da terceira componente principal (CP3), cujas correlações desta componente com os dois últimos elementos ostentam sinal negativo e com o Cd sinal positivo;
- o elemento químico Mn é o que tem mais peso sobre a componente principal (CP4).

Tabela 4.11: *Loadings*, valores das correlações entre os elementos químicos e as componentes principais e respetivos valores da variância explicada por cada componente principal na campanha de 2006.

	Loadings				Correlações			
	CP1	CP2	CP3	CP4	CP1	CP2	CP3	CP4
As	0.12	-0.55	-0.36	-0.20	0.24	-0.69	-0.41	-0.19
Cd	0.40	0.05	0.43	-0.11	0.77	0.06	0.49	-0.11
Cr	0.41	-0.06	0.27	-0.26	0.79	-0.08	0.30	-0.25
Cu	0.45	0.16	-0.05	0.27	0.86	0.19	-0.05	0.26
Fe	0.26	-0.52	-0.28	-0.09	0.50	-0.64	-0.32	-0.09
Hg	0.25	0.15	-0.44	-0.32	0.47	0.19	-0.50	-0.31
Mn	0.13	-0.43	0.15	0.74	0.24	-0.53	0.17	0.71
Pb	0.32	0.22	-0.09	0.09	0.62	0.27	-0.10	0.09
Zn	0.43	0.07	0.14	-0.05	0.82	0.09	0.16	-0.05
V	0.15	0.37	-0.54	0.37	0.29	0.45	-0.62	0.36
Valores singulares	1.92	1.24	1.15	0.97				
Proporção de variância (%)	0.37	0.15	0.13	0.09				
Proporção acumulada (%)	0.37	0.52	0.66	0.75				

De seguida, apresentam-se os planos dos *scores* da segunda versus terceira componentes principais, para as várias campanhas, representados na Figura 4.14.

Através da análise da Figura 4.14(a) (campanha de 1992), verifica-se a existência de uma elevada correlação entre os elementos Zn e Pb, sendo que a ação antropogénica é a fonte de contaminação mais evidente (Sérgio et al., 1993). Observa-se, também, que os elementos Fe, Mn e Cr, Cd estão discriminados pela segunda componente principal, que são elementos pouco tóxicos, sendo dependentes, essencialmente, da ação direta das partículas originadas por agentes erosivos e da componente geológica (Sérgio et al., 1993). Parece que a segunda componente principal está a desvincular os elementos químicos muito tóxicos (primeiro e quarto quadrante) dos elementos menos tóxicos (segundo e terceiro quadrante).

A partir do análise da Figura 4.14(b) (campanha de 1997), verifica-se que a segunda componente principal parece discriminar, em geral, um primeiro grupo de elementos químicos com origem geogénica (Mn, Fe e Cr) e um outro de elementos químicos com origem antropogénica (Pb, Zn e Cu), sendo que o metal que está mais correlacionado com a segunda componente principal é o Zn. Estas conclusões estão de acordo com Figueira (2002).

Relativamente à campanha de 2002, verifica-se através da análise da Figura 4.14(c) que no primeiro quadrante encontram-se os elementos químicos Cu, Pb e Zn que têm relação com a poluição urbana (tráfego automóvel). Por outro lado, o elemento químico Mn apresenta-se no segundo quadrante, estando relacionado com atividades ligadas à aplicação de pesticidas e à contaminação por partículas de solo. Relativamente à terceira componente principal, tem-se por um lado os elementos químicos Cu, Cd, Pb, Zn, Fe e Mn que estão relacionados com fontes de poluição antropogénica e

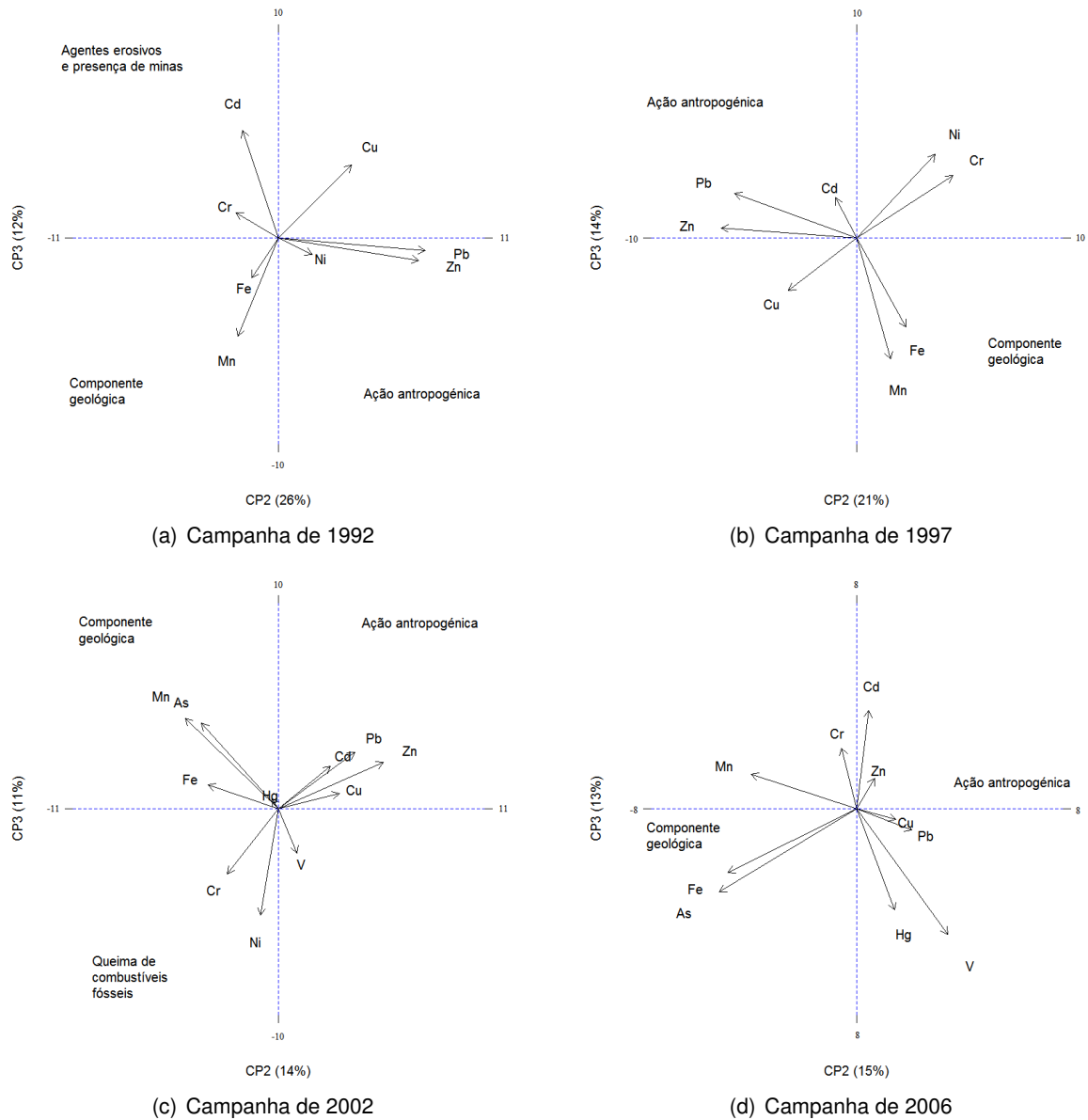


Figura 4.14: Planos dos *scores* da segunda versus terceira componentes principais determinadas em cada campanha.

presença de agentes erosivos e minas; e por outro lado, os elementos Cr e Ni, cuja associação é característica das emissões resultantes da queima de combustíveis fósseis pela indústria.

Através da análise da Figura 4.14(d), relativamente à segunda componente principal é visível por um lado o facto dos elementos Fe, Mn, As e Cr estarem presentes no segundo e terceiro quadrante, uma vez que estão associados à contaminação por partículas de solo; e por outro lado, temos os restantes elementos químicos que de uma forma geral estão associados à ação antropogénica.

Capítulo 5

Identificação e Modelação Espacial das Principais Fontes de Poluição

Como já referido, determinados organismos vivos apresentam capacidades de acumulação de elementos químicos e uma sensibilidade à poluição e, por este motivo, têm sido largamente utilizados como forma de quantificar a concentração desses poluentes no ambiente. Assim, o uso destes organismos como biomonitores, através da determinação da concentração acumulada destes elementos nos musgos, funciona como um indicador da concentração dos mesmos no ambiente, tornando-se numa metodologia que proporciona uma interpretação relevante em termos da qualidade ambiental.

Desta forma, pretende-se fazer um estudo estatístico com base na análise multivariada das concentrações dos vários elementos químicos, amostrados através de uma rede de biomonitorização em Portugal continental. O objetivo deste capítulo é a identificação das diferentes fontes de poluição que estão relacionadas com as concentrações dos elementos químicos através do conhecimento das relações lineares entre os múltiplos conjuntos de dados a que se tem acesso. Estas relações são usualmente estudadas através da Análise em Componentes Principais (ver subsecção 4.2.2), restrito a um contexto exploratório. Esta técnica permite a extração da informação essencial a partir do cálculo de novas variáveis não correlacionadas, resultando da combinação linear de todos os elementos químicos. No entanto, utilizando esta metodologia não se tem acesso aos diferentes tipos de contaminação ambiental, propondo-se definir as concentrações de um elemento químico através de uma combinação linear de todas as fontes de poluição. Assim, recorre-se a uma das metodologias da análise multivariada, designada por análise fatorial, a qual permite que estas variáveis possam ser tratadas como contínuas, resultando da separação de um número limitado de associações discretas, cujo método fornece um meio para testar se esta divisão é estatisticamente significativa.

Após a identificação das diferentes fontes de poluição, torna-se importante a aplicação de técnicas de modelação geoestatística sobre os *scores* das variáveis latentes. Deste modo, é possível a obtenção de mapas de valores preditos das diferentes fontes de poluição. Assim, consegue-se resumir a informação obtida através das concentrações dos vários elementos químicos, determinando-se, simultaneamente, os locais que padecem de determinada fonte de poluição.

5.1 O porquê da utilização da análise fatorial

Os principais objetivos dos estudos de biomonitorização são a determinação das concentrações dos vários elementos químicos em locais não amostrados, com o intuito de obter mapas de predição para cada elemento e a identificação das relações lineares existentes entre estes elementos. Para o segundo objetivo, os investigadores utilizam o método ACP, interpretando as componentes principais como as novas variáveis de síntese. No entanto, com este tipo de metodologia não se tem acesso aos diferentes tipos de contaminação ambiental, pois a ACP é uma metodologia focada, fundamentalmente, nas variáveis e nas suas contribuições. Assim, surge o interesse pela AF, uma vez que permite obter valores numéricos associados, por exemplo, ao nível da ação antropogénica.

Apesar das diferenças entre a ACP e a AF, se os valores das variâncias específicas são pequenos, os resultados de ambas as análises são muito semelhantes. Todavia, se os valores das variâncias específicas são elevados, estes irão ser absorvidos por todas as componentes principais, tanto nas componentes retidas como nas componentes rejeitadas, enquanto que a AF tem uma posição especial para estes valores. Uma vez que ambas as metodologias são aplicadas neste estudo, torna-se necessário a clarificação da disparidade entre estas duas metodologias (Chatfield and Collins, 1980):

- a técnica ACP tem como princípio a transformação das variáveis observadas em componentes principais, enquanto a AF reside na transformação das variáveis observadas a partir dos fatores;
- enquanto que a ACP é uma técnica geométrica que procura encontrar uma rotação ortogonal das variáveis, a AF baseia-se num modelo estatístico adequado aos dados (similar à regressão múltipla), dando mais ênfase à explicação da estrutura de covariância das variáveis do que às variâncias destas;
- relativamente à ACP, se o número de componentes principais a reter for aumentado, por exemplo de m para $m + 1$, as primeiras m componentes principais ficam inalteradas. Este facto não acontece na AF, uma vez que podem existir alterações substanciais em todos os fatores, caso o número de variáveis latentes a reter seja alterado;
- normalmente não existe uma relação entre as componentes principais da matriz de correlação e da matriz de covariância da amostra. No entanto, para o método de máxima verosimilhança da AF, os resultados da análise sob uma qualquer matriz são equivalentes.

A AF procura dar uma resposta à questão sobre as especificidades e as razões da estrutura dos elementos analisados, permitindo definir e avaliar a organização estrutural do sistema. Este tipo de metodologia é aplicada em diversas áreas do conhecimento, maioritariamente nas ciências sociais e com menos frequência nas ciências ambientais. As ciências ambientais lidam com sistemas caracterizados por uma variabilidade inerente (podendo ser natural, antropogénica, espaço-temporal, etc.), por uma origem multivariada e pelas relações do desenvolvimento de controle e ação segundo as leis de

probabilidade. Portanto, as relações reveladoras, bem como os seus limites e a sua hierarquia, devem ser os principais objetivos da análise de sistemas ambientais complexos (Kaplunovsky, 2005).

Esta técnica é a principal e a mais antiga técnica de análise multivariada. A ideia fundamental desta metodologia foi proposta por Spearman e por Pearson, com o objetivo de entender problemas relacionados com a psicologia educacional, na tentativa de definir inteligência (Marriot, 1974). Devido aos cálculos extremamente complexos e à falta de interpretação dos fatores, o desenvolvimento e a utilização desta técnica foram limitados durante muitos anos, sendo vista apenas como uma teoria científica. Com o desenvolvimento computacional e a permissão de diferentes rotações sobre a matriz dos *loadings*, o interesse dos investigadores por esta análise foi retomado (Menezes et al., 1978).

5.2 Identificação das principais fontes de poluição

O objetivo desta análise é identificar as relações existentes entre as variáveis latentes (correspondendo, neste caso, às diferentes fontes de poluição) com as variáveis mensuráveis, ou seja, com as concentrações dos elementos químicos nos musgos amostrados em vários locais de Portugal continental, nas diferentes campanhas.

Para obter-se tal relação, a ideia é descobrir fatores que estejam muito correlacionados com as variáveis originais. Desta forma, é necessário a estimação dos parâmetros do modelo de AFE,

$$X_j = \lambda_{j1}f_1 + \lambda_{j2}f_2 + \cdots + \lambda_{jm}f_m + \varepsilon_j.$$

A adequação desta análise aos dados realizou-se aquando da aplicação da ACP, através do teste de esfericidade de Bartlett, cujos resultados estão apresentados na Tabela 4.7 (ver subsecção 4.2.2), verificando-se a existência de correlação entre os vários elementos químicos ($p - \text{valor} < 0.05$).

De entre as muitas formas de estimar estes parâmetros, uma possibilidade é a utilização do método de mínimos quadrados ordinários, sendo que este produz resultados muito semelhantes ao método de máxima verosimilhança. Daí a utilização deste método, aplicando-se a rotação *varimax*, que se trata de um método de rotação ortogonal, com o objetivo de facilitar e viabilizar a interpretação dos fatores.

A questão que se coloca neste momento é qual o número de fatores a reter, proporcionando um modelo ótimo. Neste sentido, recorre-se à AFC que consiste num método estatístico que tem por objetivo testar se um modelo complexo, contendo os vários fatores, fornece um ajuste adequado para as covariâncias ou correlações entre as variáveis mensuráveis.

De forma a selecionar o número de fatores a reter, apresentam-se vários indicadores da qualidade do ajustamento (ver Tabela 5.1), nomeadamente o valor de prova correspondente à estatística de teste (*ET*) associada à função de discrepância de máxima verosimilhança FMV (ver equação (3.19)), supondo como hipótese nula que o número de fatores presentes no modelo é q , onde $q \in \mathbb{N}$. No

entanto, este tipo de indicador tem como requisito a normalidade dos dados, o que não acontece. Neste sentido, usa-se este valor de forma empírica e apresenta-se outros indicadores tais como RMSR, RMSEA e GFI.

Analisando a Tabela 5.1, verifica-se um bom ajustamento de um modelo com três fatores para as campanhas de 1992 e 1997. No que diz respeito às campanhas de 2002 e 2006, pode reter-se um modelo com quatro fatores. Note-se que, para este casos, um modelo com um número maior de fatores apresenta melhores indicadores de ajustamento; no entanto, os fatores deixavam de ser interpretáveis e um modelo com quatro fatores já apresenta indicadores de ajustamento razoáveis.

Tabela 5.1: Valores dos indicadores da qualidade do ajustamento dos modelos contendo 1, 2, 3, 4 e 5 fatores das várias campanhas, baseados na estimação dos parâmetros pelo método mínimos quadrados ordinários.

Campanha	Parâmetros	Número de Fatores				
		1	2	3	4	5
1992	Df ¹	20	13	7	2	
	ET	257.38	39.98	10.78	6.42	
	P-valor	< 0.001	< 0.001	0.15	0.04	
	RMSR ²	0.24	0.05	0.03	0.002	
	RMSEA ³	0.27	0.11	0.05	0.12	
	GFI ⁴	0.23	0.87	0.97	0.86	
1997	Df	20	13	7	2	
	ET	172.12	60.97	12.31	5.15	
	P-valor	< 0.001	<0.001	0.09	0.07	
	RMSR	0.16	0.08	0.03	0.02	
	RMSEA	0.21	0.15	0.06	0.10	
	GFI	0.30	0.66	0.93	0.85	
2002	Df	44	34	25	17	10
	ET	167.07	89.38	50.69	27.66	13.93
	P-valor	< 0.001	< 0.001	0.002	0.07	0.18
	RMSR	0.11	0.07	0.05	0.04	0.03
	RMSEA	0.14	0.11	0.09	0.05	0.05
	GFI	0.52	0.72	0.82	0.91	0.93
2006	Df	35	26	18	11	5
	ET	151.17	96.71	37.59	16.28	5.13
	P-valor	< 0.001	< 0.001	0.004	0.13	0.4
	RMSR	0.12	0.10	0.06	0.04	0.02
	RMSEA	0.19	0.17	0.11	0.06	0.03
	GFI	0.57	0.64	0.86	0.94	0.99

¹ Graus de liberdade associados ao modelo.

² *Root mean square residual*, representa a raiz quadrada da média das diferenças quadráticas entre os elementos de S e $\hat{\Sigma}(\theta)$. Um valor de RMSR inferior a 0.05 é geralmente considerado como uma indicação de um ajustamento razoável.

³ *Root mean square error of approximation*, indica a discrepância entre os resíduos previstos e dos observados, caso o modelo seja estimado para a população. Valores inferiores a 0.06 indicam um bom ajuste.

⁴ *Godness of fit index*, baseia-se na razão da soma do quadrado das distâncias entre a matriz de covariância observada e a matriz de covariância reproduzida pelo modelo. Valores acima de 0.9 sugerem um nível aceitável do ajustamento.

Um dos pontos fundamentais da AF é a interpretação dos fatores, obtida através da análise dos valores dos *loadings* das variáveis latentes contidos na Tabela 5.2 e nas Figuras 5.1(a), 5.1(b), 5.1(c) e 5.1(d). Desta forma, tem-se que

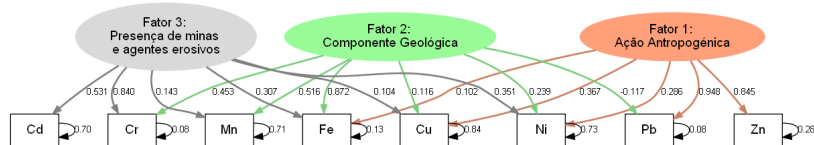
- **Campanha de 1992:** o primeiro fator envolve os elementos químicos Pb e Zn pelo que o seu rótulo pode ser *ação antropogénica*. O segundo fator pode ser designado por *componente geológica* pois envolve os elementos Fe e Mn. O terceiro fator apresenta valores mais elevados para o Cd e Cr, logo este fator pode ser legendado por *presença de minas e agentes erosivos*.
- **Campanha de 1997:** uma vez que o primeiro fator apresenta valores mais elevados para o Cr e Ni, o seu rótulo pode ser *emissões resultantes da queima de combustíveis fósseis pela indústria*. O segundo fator envolve os elementos Pb e Zn pelo que pode ser rotulado de *ação antropogénica*; e o terceiro fator pode ser rotulado por *componente geológica*, pois envolve os elementos Fe e Mn.
- **Campanha de 2002:** o primeiro fator apresenta valores mais elevados para os elementos químicos As, Cr e Fe, pelo que o seu rótulo pode ser *contaminação por partículas do solo*. O segundo fator envolve os elementos Cu, Pb, Zn e V pelo que pode ser designado por *ação antropogénica*. O terceiro e o quarto fator envolvem os elementos Ni e Mn, respetivamente, portanto podem ser legendados por *contaminação pela emissão de Ni* e de *Mn*.
- **Campanha de 2006:** o primeiro fator envolve os elementos Cd, Cr e Zn, portanto pode ser designado por *presença de minas e agentes erosivos*. O segundo fator apresenta valores mais elevados para os elementos As e Fe pelo que pode ser rotulado por *contaminação por partículas do solo*. O terceiro fator pode ser legendado por *ação antropogénica* pois envolve os elementos Pb, V e Cu. O quarto fator envolve o Hg, podendo ser rotulado por *contaminação pela emissão de Hg*.

Os valores presentes na linha identificada por ρ (Tabela 5.2) indicam que existe uma elevada correlação entre os vários fatores e os respetivos *scores*. Verifica-se, ainda, um elevado valor do *R*-quadrado o que implica uma grande proximidade entre as várias variáveis latentes e os *scores*. Os valores das variâncias específicas, representando a variabilidade específica que não é partilhada com as outras variáveis (setas a preto), estão incluídos nos pictogramas dos modelos da AF, em todas as campanhas (ver Figura 5.1). Note-se que as variáveis que têm maior peso num dado fator comum são aquelas que possuem um valor mais baixo para a variância específica e, por outro lado, as variáveis que têm um valor mais elevado para a variância específica têm pouco peso no respetivo fator comum.

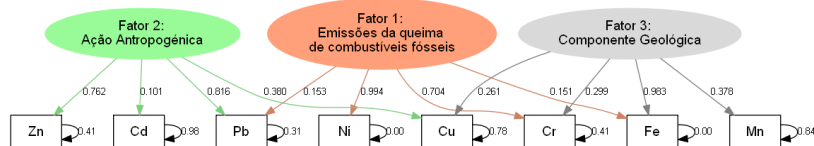
A AF, destinada a avaliar as diversas fontes de poluição, é uma ferramenta que traduz de forma fidedigna e válida estas fontes de contaminação. No entanto, a escolha do número de fatores, bem como a sua interpretação deve ser feita com o devido cuidado.

Tabela 5.2: Valores dos *loadings* das fontes de poluição subjacentes às concentrações dos elementos químicos nos musgos observadas em cada campanha.

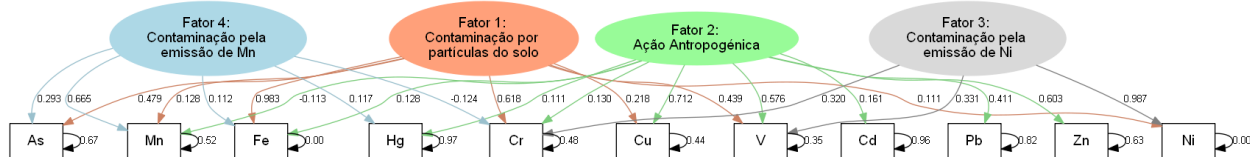
	Campanha													
	1992			1997			2002				2006			
	F1	F2	F3	F1	F2	F3	F1	F2	F3	F4	F1	F2	F3	F4
As							0.48			0.29		0.53		
Cd			0.53		0.10			0.16			0.95			
Cr		0.45	0.84	0.70		0.30	0.62	0.13	0.32	-0.12	0.80	0.24		
Cu	0.37	0.12	0.10		0.38	0.26	0.22	0.71			0.59	0.16	0.79	
Fe	0.10	0.87	0.31	0.15		0.98	0.98	0.13		0.11	0.17	0.97		0.16
Hg								0.11		0.12	0.16		0.26	0.95
Mn		0.52	0.14			0.38	0.13	-0.11		0.67	0.13	0.31	0.11	-0.17
Ni	0.29	0.24	0.35	0.99			0.11		0.99					
Pb	0.95	-0.12		0.15	0.82			0.41			0.33	0.13	0.46	
Zn	0.85				0.76			0.60			0.65	0.12	0.33	0.24
V							0.44	0.58	0.33				0.60	0.17
ρ^5	0.96	0.90	0.91	1.00	0.89	1.00	0.99	0.84	1.00	0.70	0.96	1.00	0.98	1.00
R^{26}	0.93	0.81	0.82	0.99	0.79	0.99	0.98	0.71	0.99	0.50	0.92	0.99	0.95	0.99



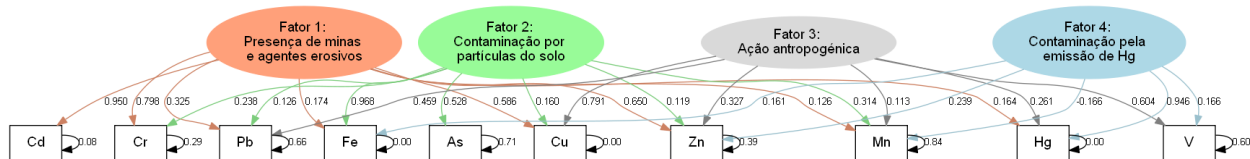
(a) Campanha de 1992



(b) Campanha de 1997



(c) Campanha de 2002



(d) Campanha de 2006

Figura 5.1: Especificação pictográfica dos modelos de análise fatorial das fontes de poluição subjacentes às concentrações dos elementos químicos nos musgos observadas em cada campanha.

⁵Correlação entre os *scores* e os fatores.

⁶Valor do *R*-quadrado. Trata-se de uma medida que quantifica o grau de proximidade entre os *scores* e os fatores.

5.3 Modelação espacial das principais fontes de poluição

Qualquer estudo de biomonitorização tem uma componente de caracterização espacial e/ou temporal. Ao estudar o impacto de uma dada fonte de contaminação ambiental (presença de minas, ação antropogénica, componente geológica, etc.), pretende-se fundamentalmente saber até onde ou em que períodos a contaminação dessa fonte ocorre, à medida que a distância a esta aumenta. Quando se avalia uma região, interessa saber qual a distribuição que as fontes de poluição apresentam e onde se localizam os valores máximos e mínimos dessa distribuição.

A geoestatística permite a caracterização da dispersão espacial das variáveis latentes, uma vez que estas apresentam uma estrutura espacial, e a interpolação espacial, utilizando os modelos que têm por base essa variabilidade estrutural. A principal vantagem na utilização da geoestatística relativamente a outros métodos de interpolação (como por exemplo, polígonos de influência, triangulação, inverso do quadrado das distâncias) consiste na incorporação da estrutura espacial da variável no processo de estimação.

O procedimento usado para a análise da variabilidade espacial das variáveis latentes é semelhante ao caso da análise dos elementos químicos (ver subsecção 4.1.2). Assim, a primeira etapa deste estudo consiste na identificação dos *outliers* presentes nas variáveis latentes, representados a vermelho nos diagramas de extremos e quartis (Anexo V). Estes valores foram eliminados na fase da estimação dos parâmetros do variograma, sendo posteriormente, incluídos na matriz de dados na fase de predição de valores para locais não amostrados. Uma vez que alguns dos processos subjacentes às fontes de poluição são considerados não estacionários, o próximo passo consiste na estimação de um modelo linear de regressão para a superfície definida pela tendência do processo estocástico espacial, considerando-se como eventuais variáveis explicativas a latitude e a longitude. Adicionalmente, estima-se os parâmetros da anisotropia. Após esta caracterização e do ajuste de um modelo teórico ao variograma empírico realizar-se-á a interpolação das variáveis latentes a toda a região em estudo, através do método de *kriging*. Efetua-se, desta forma, um estudo geoestatístico onde são aplicados modelos de *kriging* ordinário e universal (consoante as variáveis latentes).

Dado que esta metodologia já foi, detalhadamente, apresentada na subsecção 4.1.2, a apresentação das Tabelas com as estimativas dos vários parâmetros e dos gráficos dos variogramas empíricos e teóricos é feita no Anexo VI, sendo que nesta secção apenas se apresentam os mapas de predição e a respetiva interpretação.

Para qualquer fonte de poluição da campanha de 2006 verifica-se a não existência de continuidade espacial, através dos respetivos variogramas (apresentados no Anexo VI). Tal facto poderá, eventualmente, ser explicado pela dimensão reduzida da região em estudo (onde, nesta campanha, é reduzida à NUT II). Nestes casos, torna-se pouco interessante a aplicação do método *kriging* para a obtenção dos mapas de predição. Por esta razão, para esta campanha apresenta-se apenas os mapas dos valores das variáveis latentes para cada ponto amostrado.

Campanha de 1992

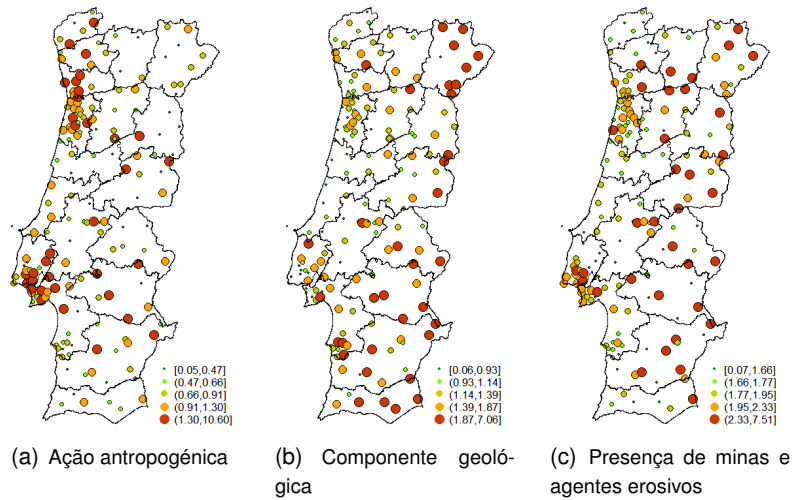


Figura 5.2: Valores dos *scores* das fontes de poluição subjacentes à campanha de 1992. Os valores são proporcionais à área do círculo.

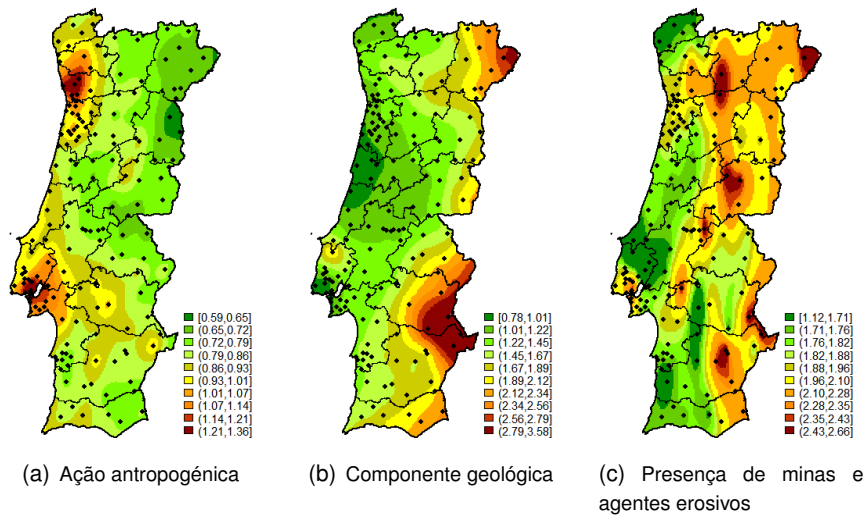


Figura 5.3: Superfícies previstas das concentrações das fontes de poluição subjacentes à campanha de 1992.

No que diz respeito à variável latente ação antropogénica (Figura 5.3(a)), verifica-se que os valores mais elevados estão contidos nas áreas urbanas tais como Lisboa, Sines, Aveiro e Porto, cujo tráfego rodoviário é mais intenso. Como resultado da influência de indústrias metalúrgicas ou do uso de inseticidas e pesticidas na agricultura, verifica-se a existência de uma mancha em Lisboa, estendendo-se a norte através do vale do Tejo. Observa-se, ainda, valores elevados desta fonte de poluição na zona sul do Tejo e Sado, possivelmente devido às atividades industriais.

Relativamente à variável latente componente geológica (Figura 5.3(b)), verifica-se que as zonas mais afetadas são o sul (junto à fronteira) e nordeste (região transmontana) de Portugal. Este fator envolve os elementos Fe e Mn, o que justifica valores elevados destes elementos nas zonas indicadas, uma vez que corresponde à forte erosão nas zonas desflorestadas pelos fogos.

No que concerne à variável latente presença de minas e agentes erosivos (Figura 5.3(c)), as con-

centrações mais elevadas deste fator encontram-se em todo o interior de Portugal (zonas rurais), o que é natural pois as concentrações do elemento químico Cr estão relacionadas com a indústria metalúrgica e extrações mineiras e a origem do elemento Cd está ligada ao uso de fertilizantes. Verifica-se, ainda, valores elevados deste fator nas minas de Jales, Mogadouro, Panasqueira, Barrancos e Neves-Corvo.

Campanha de 1997

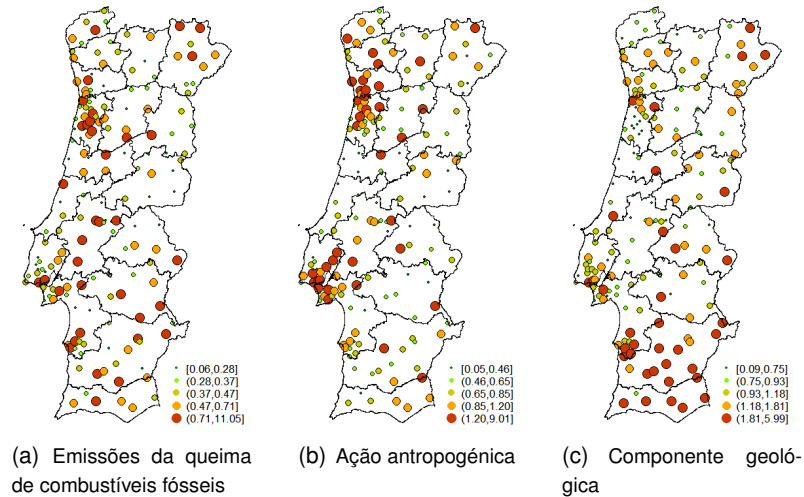


Figura 5.4: Valores dos *scores* das fontes de poluição subjacentes à campanha de 1997. Os valores são proporcionais à área do círculo.

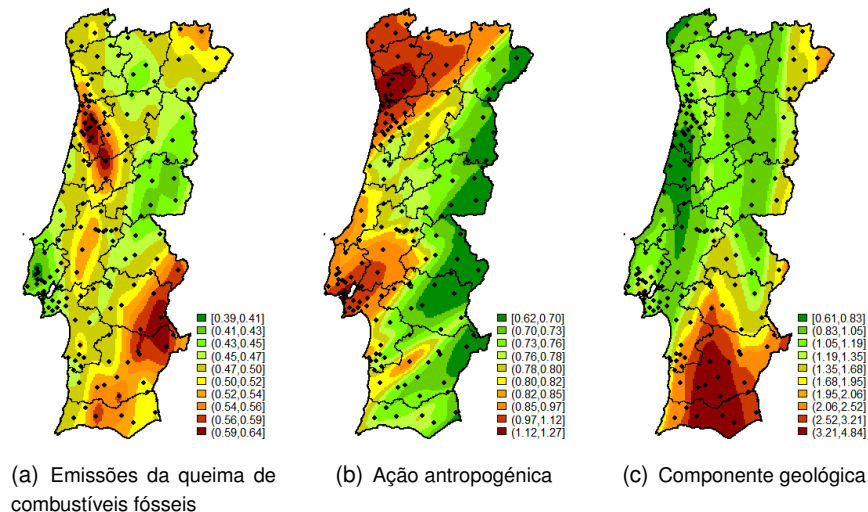


Figura 5.5: Superfícies previstas das concentrações das fontes de poluição subjacentes à campanha de 1997.

As zonas com valores mais elevados dos elementos relacionados com a variável latente emissões resultantes da queima de combustíveis fósseis pela indústria (Figura 5.5(a)) são as regiões Aveiro, Alentejo (junto à fronteira) e Algarve. A maior contaminação desta fonte acontece nas áreas mais secas e com menor cobertura vegetal do Baixo Alentejo, onde a contaminação por partículas de solo é maior. Outra área onde se verifica contaminação, de origem industrial, é na região de Aveiro.

Relativamente à variável latente ação antropogénica (Figura 5.5(b)), verifica-se que os valores

mais elevados desta fonte encontram-se na região Minho, Lisboa e Sines. Uma das causas de contaminação mais provável é o tráfego rodoviário, principal causa da ação antropogénica. A norte de Lisboa, estes valores são justificados pela presença de diversas indústrias químicas e metalúrgicas. No Minho, estes valores alastram-se a toda a região, possivelmente devido à ação contínua de soluções ricas em Cu (fungicidas) usadas no tratamento das vinhas.

No que diz respeito à variável latente presença de minas e agentes erosivos (Figura 5.5(c)), verifica-se que as zonas com valores mais elevados são o Alentejo e o Algarve. Note-se que nesta zona, nos anos em que decorreu esta campanha, existia pouco coberto vegetal e apresentava um clima seco. Assim, os solos estariam mais expostos a fatores de erosão, o que facilita a ressuspensão de partículas de solo pelo vento. É de salientar que, nestes locais, estão presentes minas situadas em Lousal, Neves-Corvo, Barrancos, Aljustrel, Nogueirinha e São Domingos.

Campanha de 2002

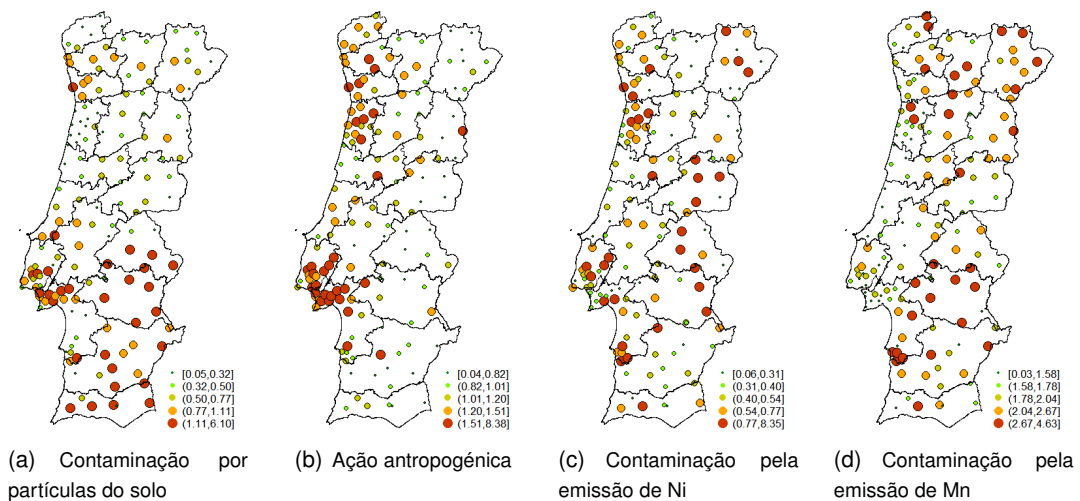


Figura 5.6: Valores dos *scores* das fontes de poluição subjacentes à campanha de 2002. Os valores são proporcionais à área do círculo.

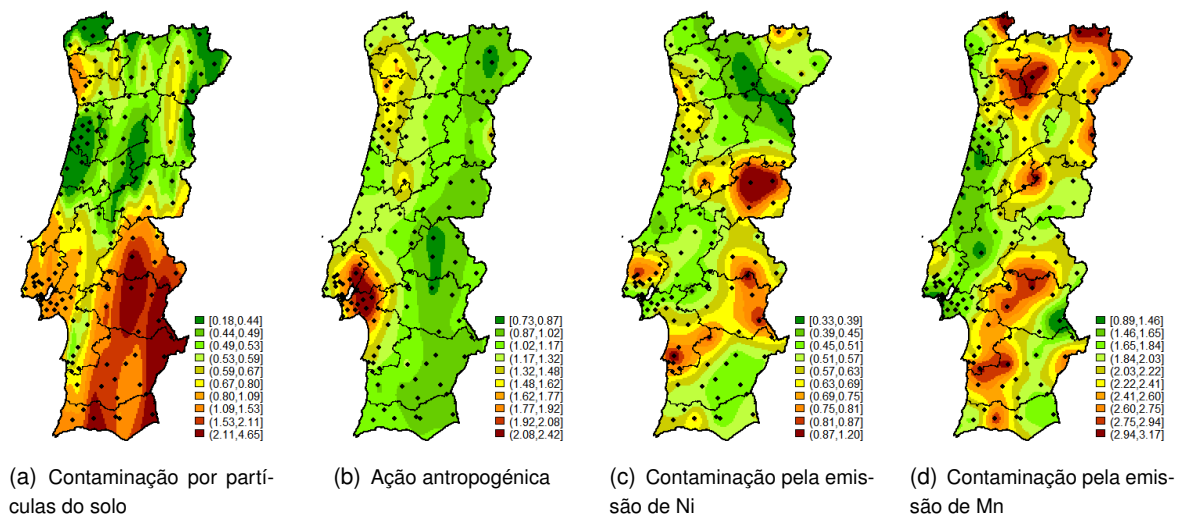


Figura 5.7: Superfícies previstas das concentrações das fontes de poluição subjacentes à campanha de 2002.

No que concerne ao (Figura 5.7(a)), a contaminação por partículas do solo ocorre no sul do país, onde existe menor densidade do coberto vegetal, estando associados a problemas de erosão. Os valores elevados desta fonte no Alentejo podem ser justificados pela presença de minas (Aljustrel, Barrancos, etc.) e pelas explorações de ouro e volfrâmio.

As zonas que apresentam valores mais elevados dos elementos relacionados com a variável latente ação antropogénica (Figura 5.7(b)) são as áreas industriais e metropolitanas do Porto/Aveiro e Lisboa, devido à atividade industrial (nomeadamente, indústrias metalúrgicas e refinarias) e emissões do tráfego automóvel. Outro local de contaminação, com origem numa central termoelétrica, é o caso da região de Sines.

Relativamente às variáveis latentes contaminação pela emissão de Ni (Figura 5.7(c)) e Mn (Figura 5.7(d)), a interpretação destes fatores será muito semelhante à que foi dada para estes dois elementos químicos. No caso do terceiro fator, verifica-se valores elevados nos distritos de Castelo Branco, Portalegre/Évora, Lisboa e Sines, cujas contaminações podem resultar de emissões industriais como refinarias, indústrias químicas e centrais termo-elétricas.

No que diz respeito ao quarto fator, os valores mais elevados encontram-se nas zonas rurais do país, correspondendo à ação da erosão existente nos solos, à componente geológica e à presença de algumas minas como Mogadouro, Jales, Barrancos, Alcáçovas e Cercal.

Campanha de 2006

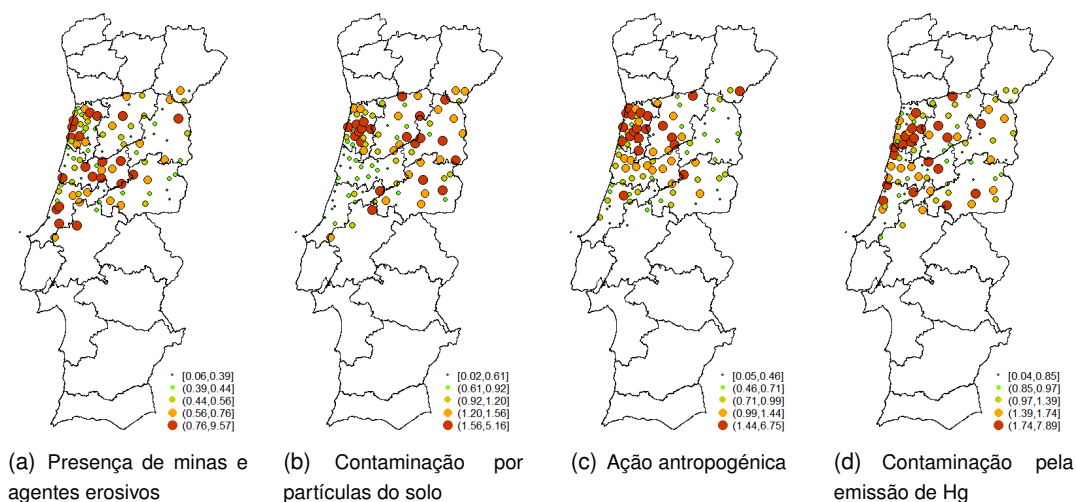


Figura 5.8: Valores dos *scores* das fontes de poluição subjacentes à campanha de 2006. Os valores são proporcionais à área do círculo.

No que diz respeito à variável latente presença de minas e agentes erosivos (Figura 5.8(a)), verifica-se que os valores elevados desta fonte de poluição parecem decrescer do litoral para o interior do país.

Relativamente à variável latente contaminação por partículas do solo (Figura 5.8(b)), parece que os valores elevados deste fator se concentram nas zonas rurais da NUT II.

A zona mais afetada pela fonte de poluição ação antropogénica (Figura 5.8(c)), parece ser Aveiro.

No que concerne à variável latente contaminação pela emissão de Hg (Figura 5.8(d)), os valores mais elevados referentes aos *scores* da contaminação por Hg parecem concentrar-se mais na zona de Aveiro.

Capítulo 6

Conclusão

A avaliação e a caracterização da deposição atmosférica dos elementos químicos em Portugal continental, utilizando organismos vivos como biomonitores, o reconhecimento das relações lineares existentes entre os vários elementos e a reconhecimento dos locais mais afetados pelas elevadas concentrações dos elementos foram os objetivos deste trabalho. No entanto, uma vez que as concentrações destes elementos podem resultar de várias fontes de contaminação ambiental, como por exemplo antropogénicas, geogénicas, geológicas, etc., tornou-se imprescindível o estudo das relações entre os vários elementos químicos e estas fontes de poluição. Desta forma, o objetivo primordial deste estudo foi a compreensão de quais as fontes de poluição que estão subjacentes a um dado elemento químico.

Para alcançar estes objetivos, foi utilizado o sistema de biomonitorização ambiental. O uso deste sistema constitui uma boa contribuição e suporte no crescente desafio de melhor compreender os processos biogeoquímicos e antropológicos da utilização e gestão ambiental, na perspetiva da utilização correta e sustentada dos recursos naturais. Este tipo de metodologia apresenta vantagens relativamente a outros métodos, permitindo a definição de uma rede de amostragem para a área em estudo ao longo das campanhas, com benefícios científicos e económicos. A biomonitorização consiste num tipo de monitorização ambiental que produz resultados com um significado bem definido para uma avaliação efetiva dos efeitos dos contaminantes no ecossistema, uma vez que se identifica, caracteriza e integra as relações entre o organismo e o ambiente. A aplicação desta metodologia permitiu a determinação das distribuições características de cada elemento químico, nas diferentes campanhas, e a identificação das principais causas de contaminação de cada elemento.

De forma a caracterizar a continuidade espacial do processo subjacente a um dado elemento, recorreu-se à construção de variogramas. No entanto, constatou-se que alguns destes processos não verificavam a hipótese de estacionaridade e isotropia, sendo necessário a estimação de alguns parâmetros relativos à componente da tendência espacial e à anisotropia. Os mapas produzidos pelo método *kriging* permitiram a identificação das áreas de maior deposição dos elementos, exibindo a distribuição média dos elementos. Este é um método de interpolação estocástica de inferência espacial, onde as concentrações dos elementos nos musgos são consideradas realizações de processos aleatórios, sendo permitida a avaliação da qualidade das estimativas obtidas, através do desvio pa-

drão da estimação. Verificou-se que os elementos que têm, na generalidade, fontes de contaminação geogénicas são o Fe, Mn e Cr, enquanto que o Cu, Pb e Zn têm, no geral, fontes antropogénicas.

Um outro ponto importante foi estabelecer relações entre os níveis dos elementos determinados nos musgos e as fontes de poluição, investigadas com base nos métodos de análise multivariada de dados. Os elementos químicos analisados apresentam associações que derivam naturalmente de fontes de contaminação comuns, porém, verificou-se que estas associações podem expressar-se de maneira diferente quando se toma em linha de conta não só a sua variabilidade, mas também a covariância entre os vários elementos. Este facto é a principal diferença entre os métodos ACP e AF, ambos aplicados neste estudo. Com a metodologia ACP não se tinha acesso às diferentes fontes de contaminação ambiental, que levam à deposição destes elementos nos musgos. Todavia, a AF permitiu a obtenção das concentrações de um dado elemento químico à custa de uma combinação linear das várias fontes de poluição e não uma nova variável que resulta da combinação de todos os elementos químicos. Desta forma, procedeu-se a uma análise fatorial exploratória e confirmatória. Os resultados finais desta metodologia são coerentes com o que se conhece do fenómeno em estudo e com o que se conhece da região onde este fenómeno está a ser observado, uma vez que as distribuições dos mapas de predição dos *scores* das variáveis latentes correspondem àquilo que se conhece acerca das fontes de poluição.

Em termos de desenvolvimentos futuros, perspetiva-se

- a integração de outras informações cartográficas nos modelos desenvolvidos, uma vez que ao nível das metodologias geoestatísticas aplicadas sentiu-se que são necessários desenvolvimentos na forma de descrever os fenómenos pelos modelos construídos; nomeadamente, na integração de outras variáveis explicativas (covariáveis), tais como precipitação, distância a um nó rodoviário, velocidade do vento, distância a uma mina, distância à costa, etc., na aplicação do método *kriging*;
- a utilização de modelos de simulação estocástica de processos espaciais, dando um conjunto de imagens equiprováveis do fenómeno físico com a mesma variabilidade e continuidade espacial conforme revelado pelos dados experimentais;
- a análise das relações entre os diferentes cenários temporais com a capacidade de retenção do biomonitor, com o objetivo de determinar de uma forma mais correta o comportamento químico dos biomonitores relativamente a este tipo de cenários.

Bibliografia

- Anderberg, S., Prieler, S., Olendrzynski, K., and Bruyn, S. (2000). *Industrial Metabolism, Heavy Metal Pollution and Environmental Transition in Central Europe*. United Nations University Press, USA.
- Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*. Wiley, USA, 3rd edition.
- Bartholomew, D. and Knott, M. (1999). *Kendall's Library of Statistics 7: Latent Variable Models and Factor Analysis*. Hodder Arnold, UK, 2nd edition.
- Berg, T. and Steinnes, E. (1997). Recent trends in atmospheric deposition of trace elements in norway as evident from the 1995 moss survey. *The Science of the Total Environment*, 208:197–202.
- Bowman, A. and Crujeiras, R. (2013). Inference for variograms. *Computational Statistics and Data Analysis*, 66:19–31.
- Carvalho, M. and Natário, I. (2008). *Análise de Dados Espaciais*. Congresso Anual, Sociedade Portuguesa de Estatística, Vila Real.
- Chatfield, C. and Collins, A. J. (1980). *Introduction to Multivariate Analysis*. Chapman and Hall, Cambridge.
- Crawley, M. (2007). *The R book*. John Wiley and Sons, UK.
- Cressie, N. (1993). *Statistics for Spatial Data*. Wiley, USA.
- Diem, J. and Comrie, A. (2002). Predictive mapping of air pollution involving sparse spatial observations. *Environmental Pollution*, 119(1):99–117.
- Diggle, P. and Ribeiro, P. (2007). *Model-based Geostatistics*. Springer, USA.
- Everitt, B. (1984). *An Introduction to Latent Variable Models*. Chapman and Hall, UK.
- Everitt, B. and Dunn, G. (2010). *Applied Multivariate Data Analysis*. Wiley, USA, 2nd edition.
- Everitt, B. and Hothorn, T. (2011). *An Introduction to Applied Multivariate Analysis with R*. Springer, USA.
- Everitt, B., Landau, S., Leese, M., and Stahl, D. (2011). *Cluster Analysis*. John Wiley and Sons, UK, 5th edition.
- Faus-Kessler, T., Dietl, C., Tristchler, J., and Peichel, L. (2000). Correlation patterns of metals in the epiphytic moss hypnum cupressiform in bavaria. *Atmospheric Environment*, 35(2):427–439.
- Fergusson, J. (1990). *The heavy elements: chemistry, environmental impact and health effects*. Pergamon Press, New Zealand.
- Fernández, J., Rey, A., and Carballeira, A. (2000). An extend study of heavy metal deposition in galicia (nw spain) based on moss analysis. *The Science of the Total Environment*, 254:31–44.

- Fernández, J. A., Real, C., Couto, J. A., Aboal, J. R., and Carballeira, A. (2005). The effect of sampling design on extensive bryomonitoring surveys of air pollution. *The Science of Total Environment*, 337(1-3):11–21.
- Ferreira, M. (2004). *Atlas Geoquímico dos Solos de Portugal Continental*. Tese de Doutoramento, Departamento de Geociências, Universidade de Aveiro, Aveiro.
- Figueira, R. (2002). *Desenvolvimento de um Sistema de Biomonitorização Ambiental*. Tese de Doutoramento, Faculdade de Ciências da Universidade de Lisboa, Lisboa.
- Figueira, R., Ferreira, J., Sérgio, C., Ramalho, C., and Sousa, A. (2004). *Distribuição da deposição atmosférica de metais em Portugal analisada através de musgos biomonitores*. 8a Conferência Nacional do Ambiente, Lisboa.
- Figueira, R., Sérgio, C., and Sousa, A. (2002). Distribution of trace metals in moss biomonitors and assessment of contamination sources in Portugal. *Environmental Pollution*, 118:153–163.
- Gerdol, R., Bragazza, L., Marchesini, R., Alber, R., Bonetti, L., Lorenzoni, G., Achilli, M., Achilli, M., Buffoni, A., De Marco, N., Franchi, M., Pison, S., Giaquinta, S., Palmieri, F., and Spezzano, P. (2000). Monitoring of heavy metal deposition in northern Italy by moss analysis. *Environment Pollution*, 108:201–208.
- Goovaerts, P. (1997). *Geostatistics for Natural Resources Evaluation*. Oxford University Press, USA.
- Gordon, A. (1987). A review of hierarchical classification. *Journal of the Royal Statistical Society*.
- Gordon, A. (1999). *Classification*. Chapman and Hall, UK, 2nd edition.
- Grodzinska, K., Szarek-Lukaszewska, G., and Godsik, B. (1999). Survey of heavy metal deposition in Poland using mosses as indicators. *The Science of the Total Environment*, 229:41–51.
- Harrison, R. (2001). *Pollution: Causes, Effects and Control*. The Royal Society of Chemistry, UK, 4th edition.
- Hill, M. (2010). *Understanding Environmental Pollution*. Cambridge University Press, UK, 3rd edition.
- Jobson, J. D. (1992). *Applied Multivariate Data Analysis: Categorical and Multivariate Methods*. Springer, USA.
- Johnson, R. and Wichern, D. (2007). *Applied Multivariate Statistical Analysis*. Prentice Hall, USA, 6th edition.
- Jolliffe, I. T. (2002). *Principal Component Analysis*. Springer, USA, 2nd edition.
- Kaplunovsky, A. (2005). Factor analysis in environmental studies. *Journal of Science and Engineering B*.
- Kyriakidis, P. and Journel, A. (1999). Geostatistical space-time models: a review. *Mathematical Geology*.
- Lloyd, C. and Atkinson, P. (2004). Increased accuracy of geostatistical prediction of nitrogen dioxide in the United Kingdom with secondary data. *Applied Earth Observation and Geoinformation*, 5(4):293–305.
- Marriot, F. (1974). *The interpretation of multiple observations*. Academic Press, New York.
- Martins, A. (2011). *Interpolação geoestatística espaço-temporal de Cu em musgos usando interpolação da distribuição local*. Lisboa.
- McGranahan, G. and Murray, F. (2003). *Air Pollution and Health in Rapidly Developing Countries*. Earthscan, UK and USA.

- Menezes, A., Faissol, S., and Ferreira, M. (1978). *Tendências atuais na geografia urbano-regional: teorização e quantificação*. IBGE, Instituto Brasileiro de Geografia e Estatística, Brasil.
- Menezes, R. (2007). *Modelação Geoestatística da Deposição de Poluentes Atmosféricos em Sistemas de Informação Geográfica, para a Região de Aveiro: Integração de dados de Biomonitorização e de Modelação Física*. Tese de Mestrado, Instituto Superior Técnico, Universidade Técnica de Lisboa, Lisboa.
- Militino, A., Ugarte, M., and Ibáñez, B. (2008). Longitudinal analysis of spatially correlated data. *Stochastic Environmental Research and Risk Assessment*, 22(1):49–57.
- Morrison, D. (2005). *Multivariate Statistical Methods*. McGraw-Hill, New York, 4th edition.
- R Team, D. (2010). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.
- Reis, A., Patinha, C., Silva, E., Sousa, A., Figueira, R., Sérgio, C., and Novais, V. (2010). Assessment of human exposure to environmental heavy metals in soils and bryophytes of the central region of Portugal. *Environmental Health Research*, 20(2):87–113.
- Reis, A., Silva, E., Sousa, A., Patinha, C., and Fonseca, E. (2007). Spatial patterns of dispersion and pollution sources for arsenic at Iousal mine, Portugal. *Environmental Health Research*, 17(5):335–349.
- Reis, E. (2001). *Estatística Multivariada Aplicada*. Edições Sílabo, Lisboa, 2a edition.
- Rhoades, F. (1999). *A Review of Lichen and Bryophyte Elemental Content Literature with Reference to Pacific Northwest Species*. United States Department of Agriculture, Forest Service, Bellingham.
- Ribeiro, R. and Mingoti, S. (2006). *Apresentando um programa computacional em R e geoR para análise espaço-temporal via geoestatística*. XIII SIMPEP - Bauru, Brasil.
- Sérgio, C., Sim-Sim, M., and Figueira, R. (1993). *Quantificação da deposição de metais pesados em Portugal, através da análise de briófitos*. Direcção Geral da Qualidade do Ambiente, Lisboa.
- Soares, A. (2000). *Geoestatística para as Ciências da Terra e do Ambiente*. IST Press, Lisboa, 2a edition.
- Stein, M. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*. Springer, USA.
- Szczepaniak, K. and Biziuk, M. (2003). Aspects of the biomonitoring studies using mosses and lichens as indicators of metal pollution. *Environmental Research*, 93(3):221–230.
- Timm, N. (2002). *Applied Multivariate Analysis*. Springer, USA.
- Tobler, W. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*.
- Wackernagel, H. (2003). *Multivariate Geostatistics: An Introduction with Applications*. Springer, Germany, 3rd edition.
- Whitfield, J. (2001). Vital signs. *Nature*, 411:989–990.

Anexo I

Distritos de Portugal continental



Figura 1: Mapa de Portugal continental, referenciando as regiões contidas na NUT II, bem como os distritos de Portugal.

Anexo II

Valores das concentrações dos elementos químicos nos musgos

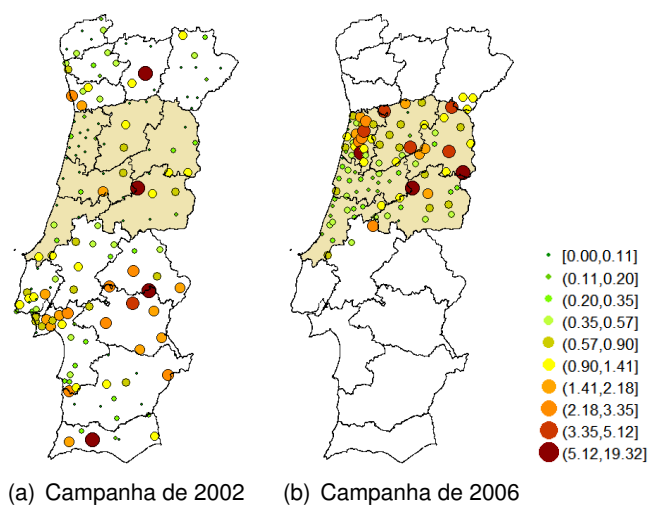


Figura 2: Valores das concentrações de As nos musgos observadas em cada campanha. Os valores das concentrações são proporcionais à área do círculo.

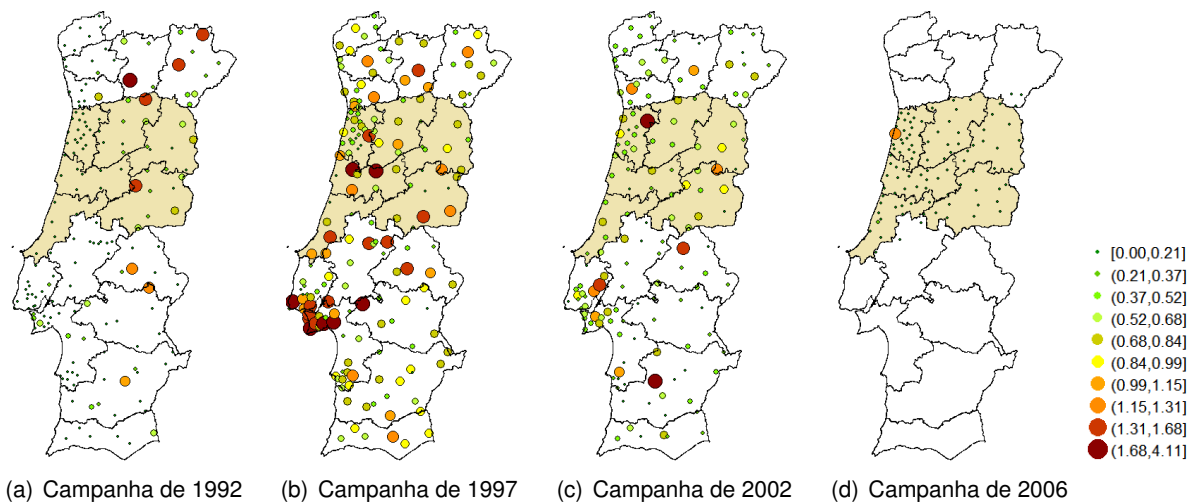


Figura 3: Valores das concentrações de Cd nos musgos observadas em cada campanha. Os valores das concentrações são proporcionais à área do círculo.

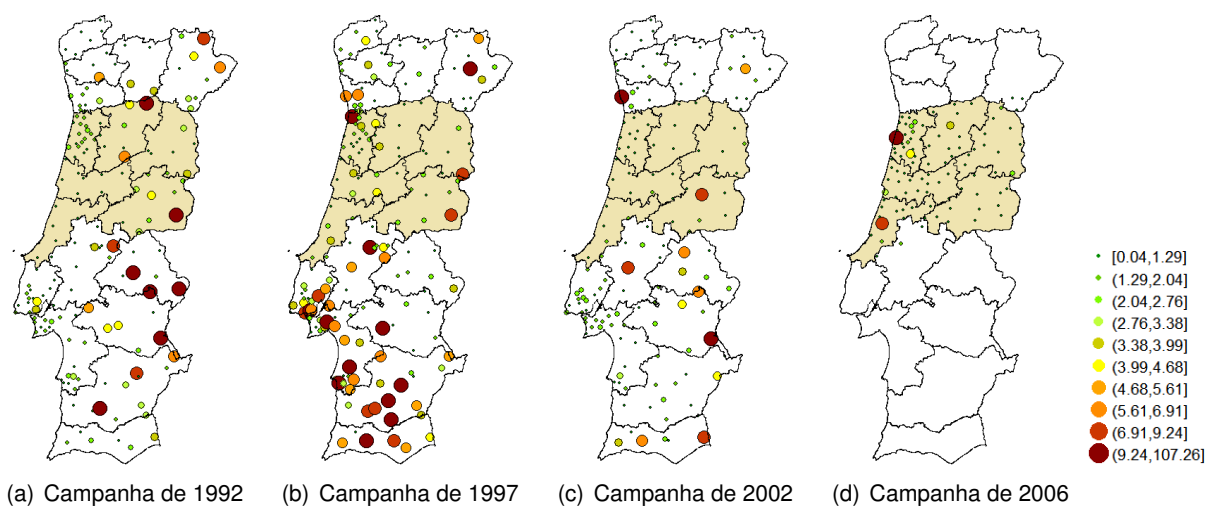


Figura 4: Valores das concentrações de Cr nos musgos observadas em cada campanha. Os valores das concentrações são proporcionais à área do círculo.

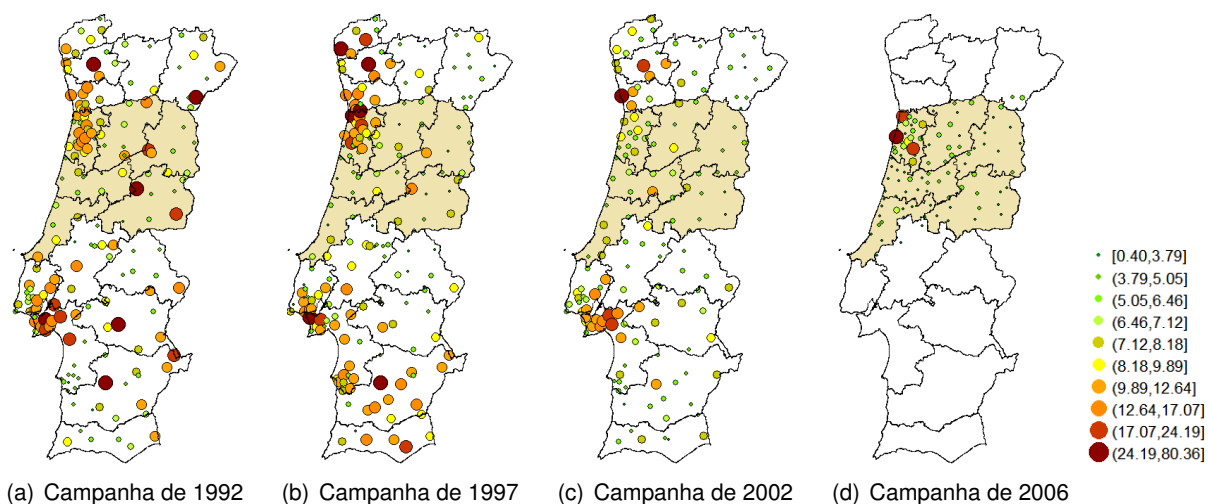


Figura 5: Valores das concentrações de Cu nos musgos observadas em cada campanha. Os valores das concentrações são proporcionais à área do círculo.

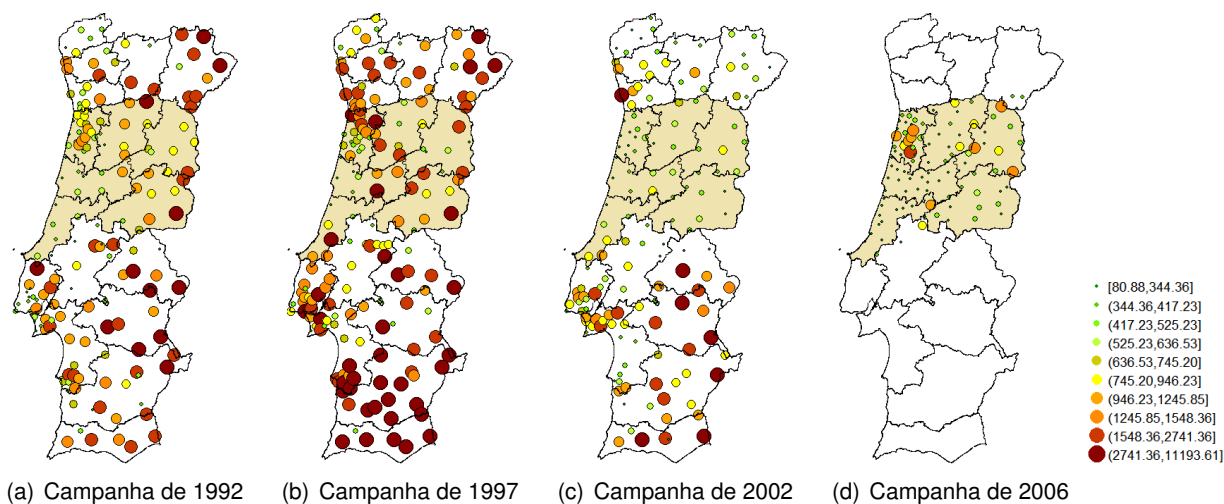


Figura 6: Valores das concentrações de Fe nos musgos observadas em cada campanha. Os valores das concentrações são proporcionais à área do círculo.

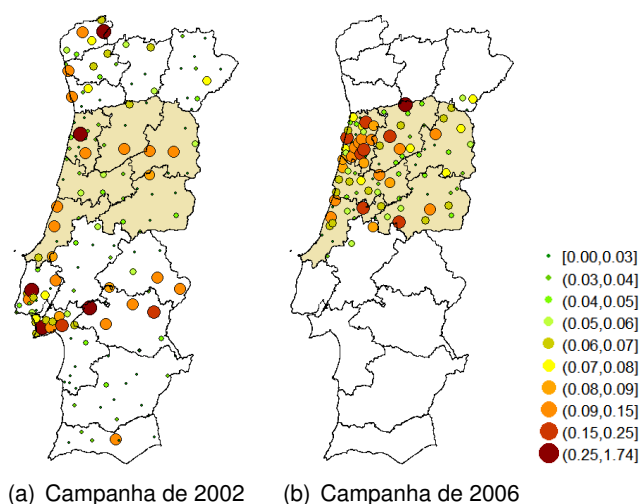


Figura 7: Valores das concentrações de Hg nos musgos observadas em cada campanha. Os valores das concentrações são proporcionais à área do círculo.

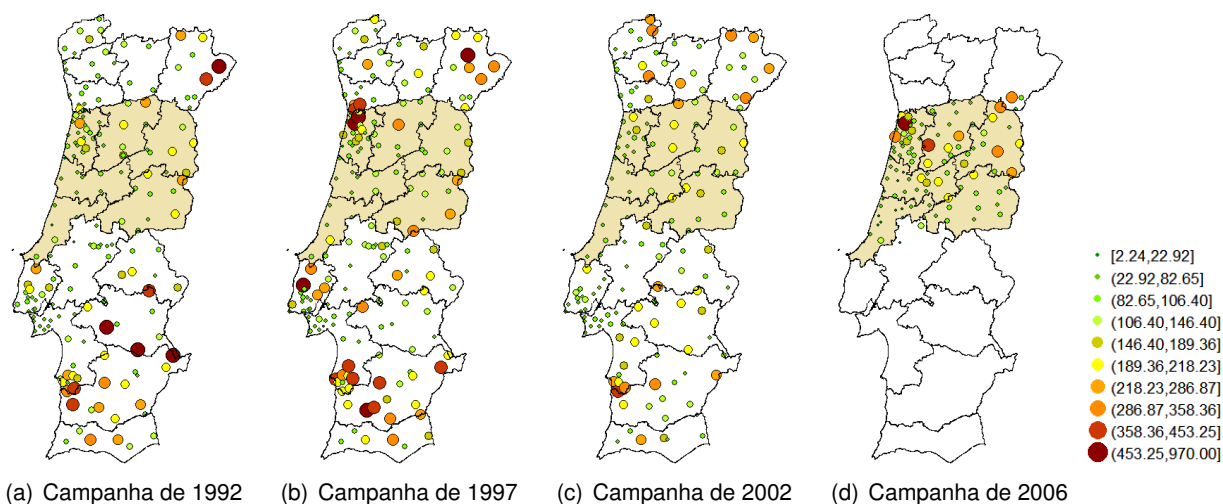


Figura 8: Valores das concentrações de Mn nos musgos observadas em cada campanha. Os valores das concentrações são proporcionais à área do círculo.

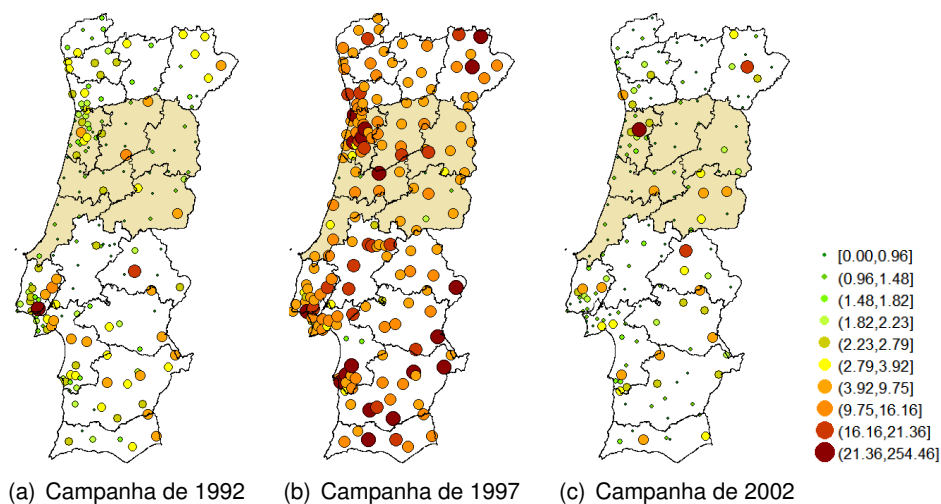


Figura 9: Valores das concentrações de Ni nos musgos observadas em cada campanha. Os valores das concentrações são proporcionais à área do círculo.

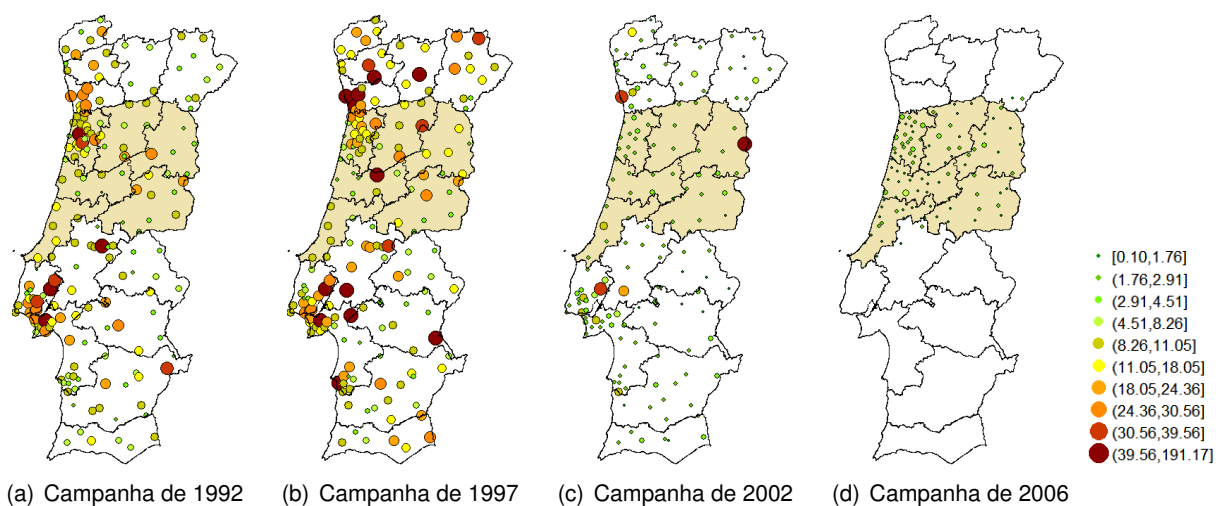


Figura 10: Valores das concentrações de Pb nos musgos observadas em cada campanha. Os valores das concentrações são proporcionais à área do círculo.

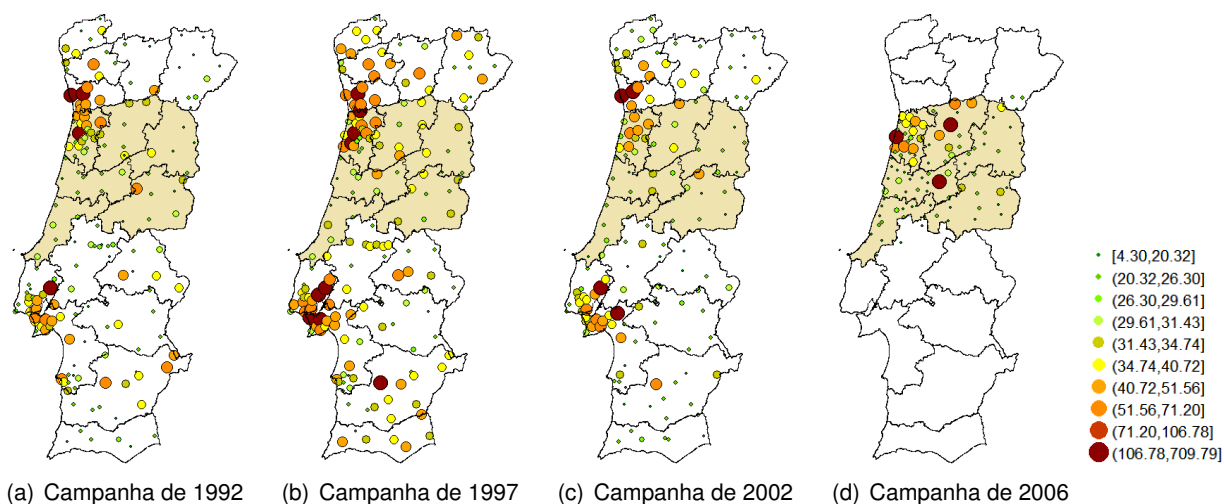


Figura 11: Valores das concentrações de Zn nos musgos observadas em cada campanha. Os valores das concentrações são proporcionais à área do círculo.

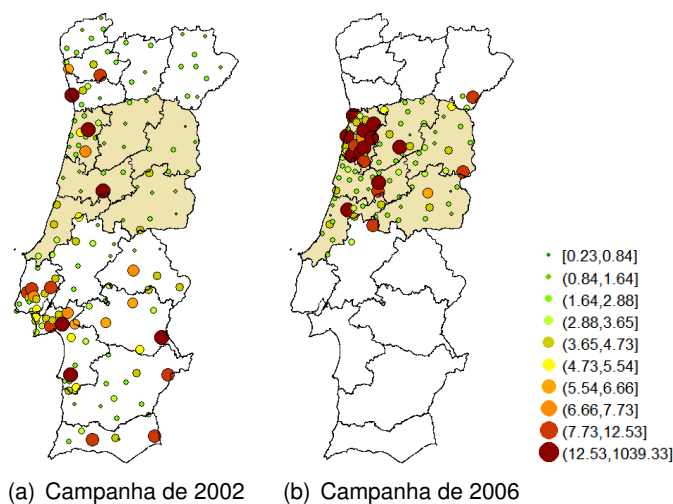
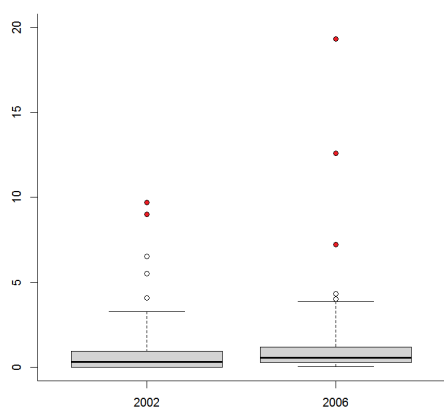


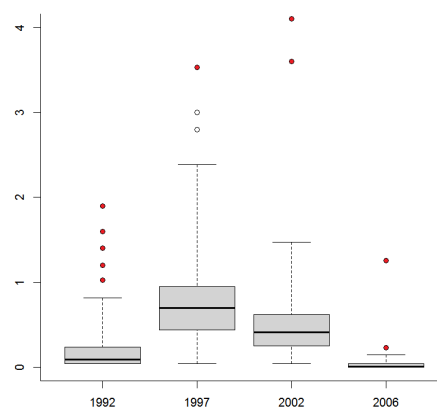
Figura 12: Valores das concentrações de V nos musgos observadas em cada campanha. Os valores das concentrações são proporcionais à área do círculo.

Anexo III

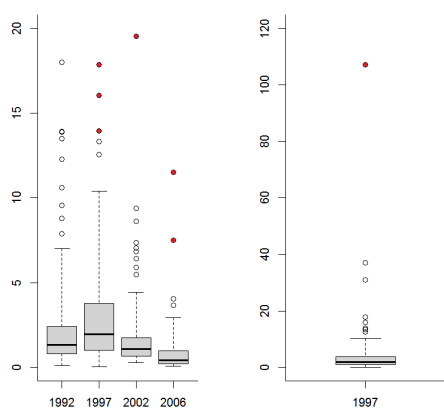
Diagramas de extremos e quartis dos elementos químicos



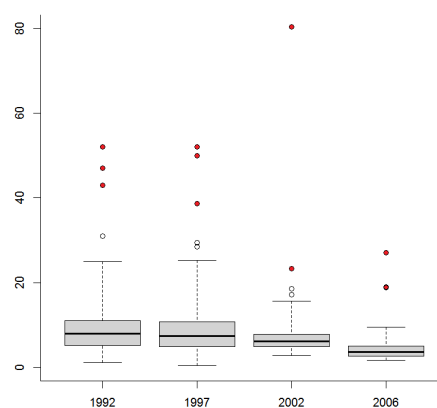
(a) Arsénio



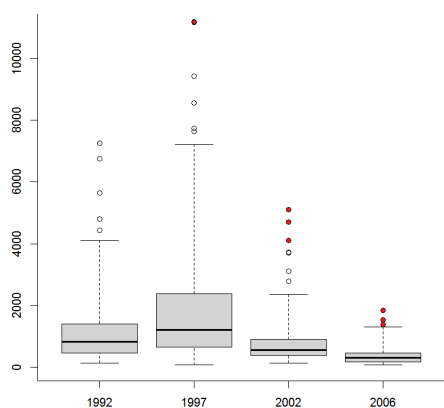
(b) Cádmio



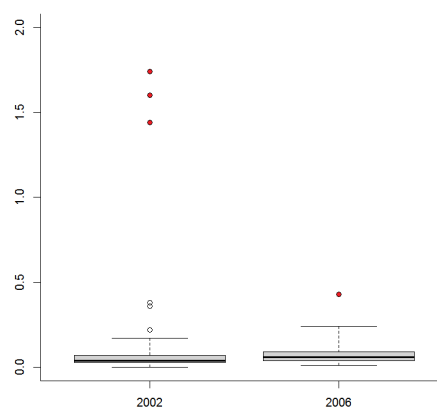
(c) Crômio



(d) Cobre



(e) Ferro



(f) Mercúrio

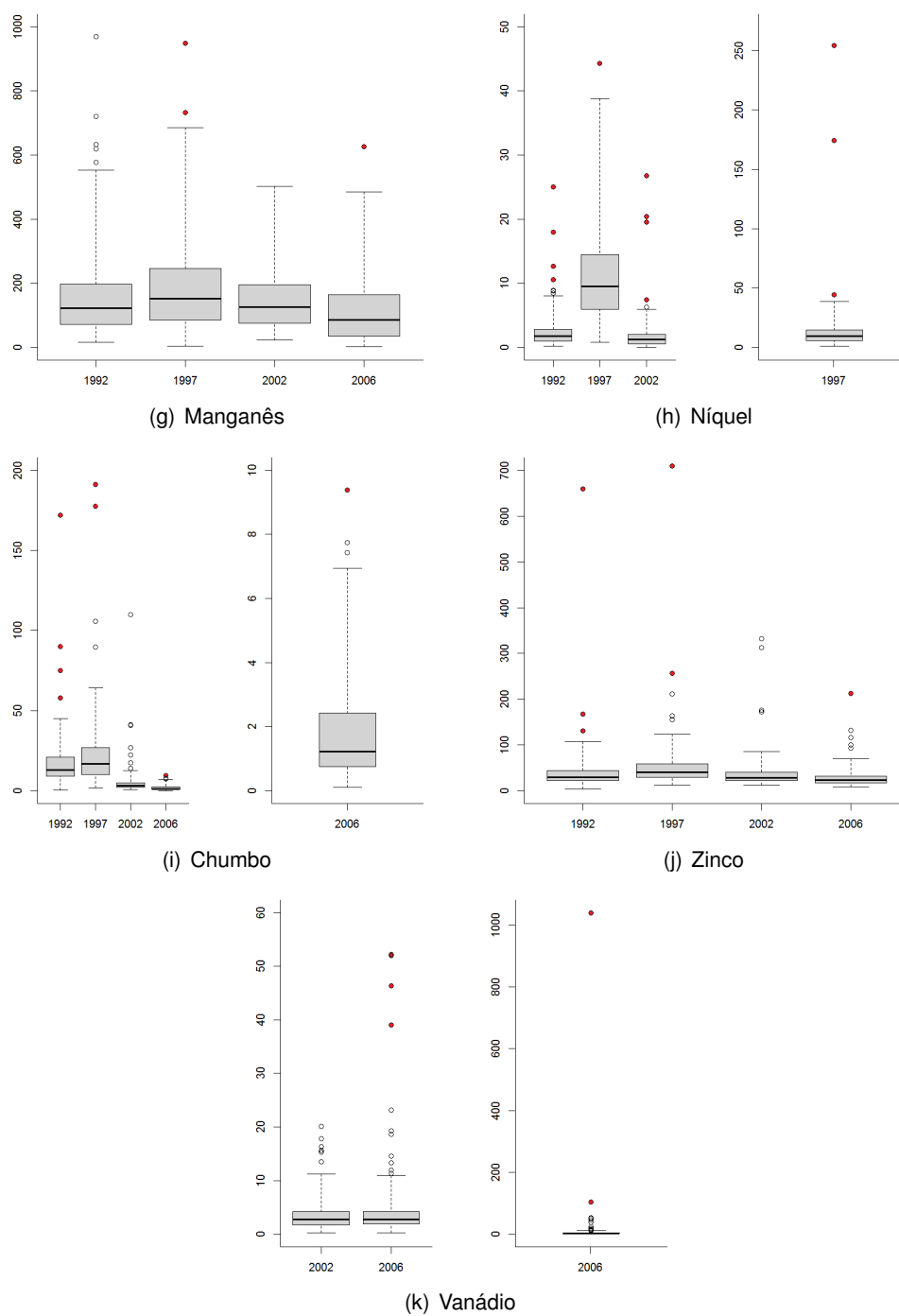


Figura 13: Diagramas de extremos e quartis das concentrações dos vários elementos químicos nos musgos observadas em cada campanha.

Anexo IV

Análise geoestatística dos elementos químicos

Cádmio

Tabela 1: Estimativas dos parâmetros envolvidos nas estruturas do valor médio e da anisotropia das concentrações de Cd nos musgos observadas em cada campanha.

Campanha	Tendência						Anisotropia	
	$\hat{\beta}_0$	D.p.	$\hat{\beta}_1$	D.p.	$\hat{\beta}_2$	D.p.	$\hat{\psi}_A$	$\hat{\psi}_R$
1992	0.17	0.01	0.001	0.000	0.000	0.000	2.51e-09	1.00
1997	0.76	0.04	-	-	-	-	0.00	3.61
2002	0.45	0.02	-	-	-	-	0.36	12.22
2006	0.04	0.01	-8.58e-05	7.36e-05	-1.46e-04	7.63e-05	1.60e-08	1.00

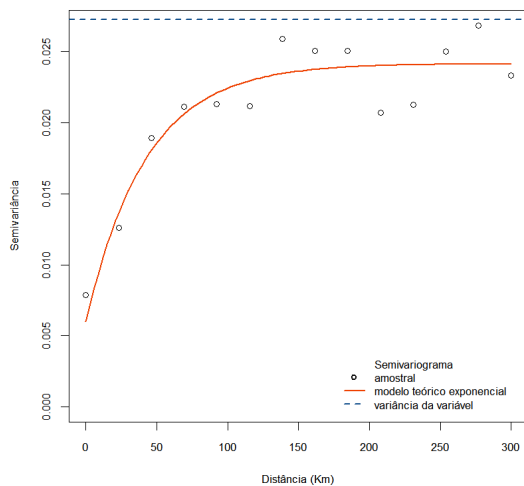
Tabela 2: Valores dos variogramas empíricos das concentrações de Cd nos musgos observadas em cada campanha, onde $|N(u)|$ representa o cardinal das diferentes classes de distâncias.

Campanha							
1992				1997			
u	$ N(u) $	$\hat{\gamma}(u)$	D.p.	u	$ N(u) $	$\hat{\gamma}(u)$	D.p.
0.00	92.00	0.01	0.02	0.00	99.00	0.17	0.41
23.08	607.00	0.01	0.03	12.50	682.00	0.28	0.56
46.15	732.00	0.02	0.04	25.00	930.00	0.25	0.47
69.23	844.00	0.02	0.04	37.50	1108.00	0.22	0.40
92.31	982.00	0.02	0.04	50.00	1235.00	0.21	0.40
115.38	1136.00	0.02	0.04	62.50	1427.00	0.23	0.43
138.46	1123.00	0.03	0.04	75.00	1526.00	0.24	0.45
161.54	1052.00	0.03	0.04	87.50	1654.00	0.24	0.46
184.62	911.00	0.03	0.04	100.00	1471.00	0.25	0.52
207.69	806.00	0.02	0.03	112.50	1290.00	0.20	0.44
230.77	788.00	0.02	0.04	125.00	1128.00	0.21	0.42
253.85	735.00	0.03	0.04	137.50	912.00	0.20	0.44
276.92	571.00	0.03	0.04	150.00	336.00	0.21	0.46
300.00	287.00	0.02	0.04				

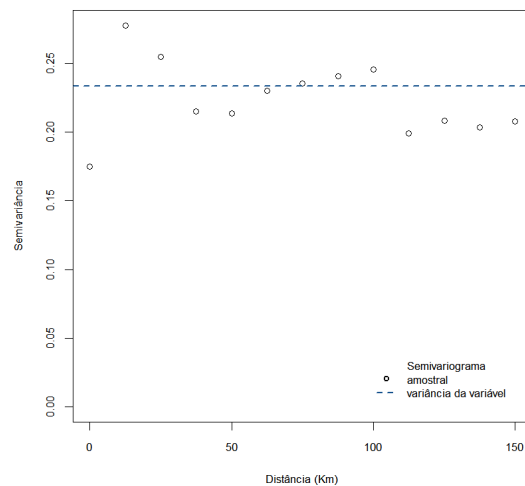
Campanha							
2002				2006			
u	$ N(u) $	$\hat{\gamma}(u)$	D.p.	u	$ N(u) $	$\hat{\gamma}(u)$	D.p.
0.00	156.00	0.08	0.11	0.00	30.00	0.00088	0.00235
12.50	939.00	0.07	0.12	20.00	402.00	0.00113	0.00203
25.00	1267.00	0.09	0.15	40.00	583.00	0.00124	0.00218
37.50	1218.00	0.08	0.13	60.00	722.00	0.00126	0.00217
50.00	1060.00	0.08	0.13	80.00	725.00	0.00110	0.00194
62.50	918.00	0.07	0.12	100.00	656.00	0.00121	0.00213
75.00	856.00	0.08	0.13	120.00	555.00	0.00106	0.00187
87.50	777.00	0.08	0.12	140.00	447.00	0.00106	0.00195
100.00	721.00	0.08	0.12	160.00	242.00	0.00137	0.00211
112.50	660.00	0.10	0.16	180.00	104.00	0.00132	0.00207
125.00	513.00	0.09	0.13	200.00	53.00	0.00138	0.00192
137.50	430.00	0.08	0.13	220.00	13.00	0.00105	0.00156
150.00	195.00	0.06	0.10				

Tabela 3: Estimativas dos parâmetros envolvidos nas estruturas da associação espacial dos modelos adotados para as concentrações de Cd nos musgos observadas em cada campanha.

Campanha	Modelo	Parâmetros					
		EM	MEEQ	SQR	$\hat{\tau}^2$	$\hat{\sigma}^2$	$\hat{\phi}$
1992	Exponencial	0.0003	1.30	0.04	0.01	0.02	42.40
	Esférico	-0.0024	2.79	0.05	0.00	0.02	69.10
	Gaussiano	0.0006	1.24	0.04	0.01	0.01	68.86
1997	-	-	-	-	-	-	-
2002	-	-	-	-	-	-	-
2006	Exponencial	0.0001	0.00008	1.11	0.00080	0.00048	21.32
	Esférico	0.0001	1.08	0.00006	0.00085	0.00040	45.23
	Gaussiano	3.60e-05	1.35	0.00010	0.00080	0.00040	42.56



(a) Campanha de 1992



(b) Campanha de 1997

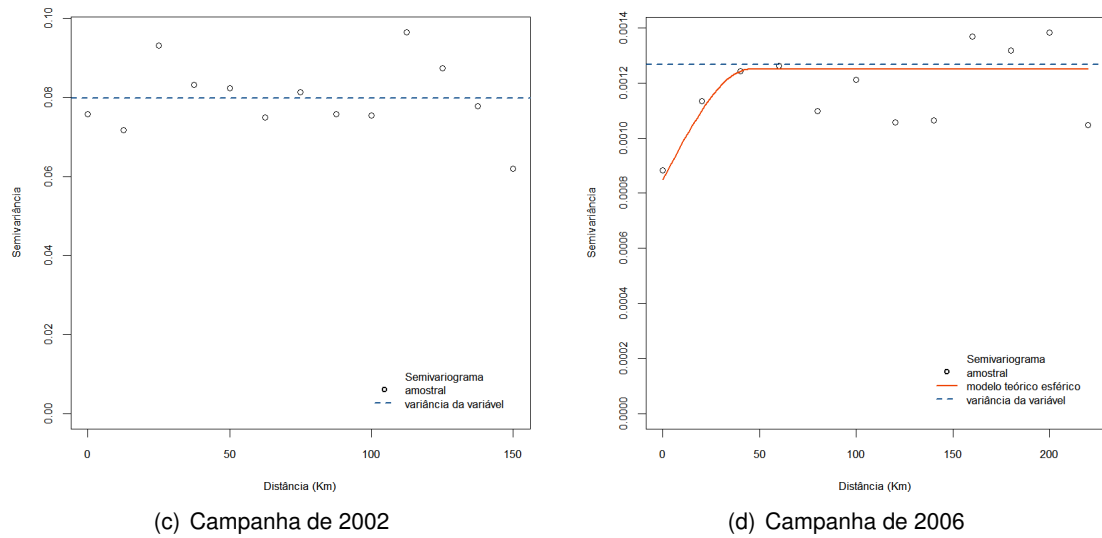


Figura 14: Variogramas amostrais das concentrações de Cd nos musgos observadas em cada campanha e respetivos modelos teóricos, estimados pelo método de mínimos quadrados.

Crómio

Tabela 4: Estimativas dos parâmetros envolvidos nas estruturas do valor médio e da anisotropia das concentrações de Cr nos musgos observadas em cada campanha.

Campanha	Tendência						Anisotropia	
	$\hat{\beta}_0$	D.p.	$\hat{\beta}_1$	D.p.	$\hat{\beta}_2$	D.p.	$\hat{\psi}_A$	$\hat{\psi}_R$
1992	2.44	0.19	0.02	0.003	-0.003	0.001	7.47e-06	1.00
1997	2.96	0.22	-0.001	0.004	-0.008	0.002	0.00	7.83
2002	1.48	0.10	0.003	0.001	-0.003	0.001	0.23	2.66
2006	0.47	0.15	-	-	-	-	0.15	9.28

Tabela 5: Valores dos variogramas empíricos das concentrações de Cr nos musgos observadas em cada campanha, onde $|N(u)|$ representa o cardinal das diferentes classes de distâncias.

Campanha							
1992				1997			
u	$ N(u) $	$\hat{\gamma}(u)$	D.p.	u	$ N(u) $	$\hat{\gamma}(u)$	D.p.
0.00	70.00	1.77	6.42	0.00	58.00	7.33	16.10
20.83	521.00	2.75	12.43	6.43	417.00	8.08	16.99
41.67	723.00	3.44	11.46	12.86	559.00	7.37	17.81
62.50	796.00	5.75	17.47	19.29	683.00	6.64	16.07
83.33	959.00	5.59	15.92	25.71	786.00	7.20	17.08
104.17	1062.00	6.84	18.65	32.14	893.00	7.59	17.03
125.00	1189.00	6.80	17.78	38.57	901.00	9.48	21.66
145.83	1094.00	8.21	19.94	45.00	1005.00	9.24	20.24
166.67	1054.00	7.52	18.72	51.43	1004.00	9.80	19.70
187.50	920.00	8.27	19.55	57.86	963.00	9.91	19.04
208.33	829.00	7.31	18.63	64.29	843.00	8.63	17.88

Campanha							
1992				1997			
u	$ N(u) $	$\hat{\gamma}(u)$	D.p.	u	$ N(u) $	$\hat{\gamma}(u)$	D.p.
229.17	763.00	6.62	16.51	70.71	752.00	8.32	15.74
250.00	385.00	5.48	14.56	77.14	676.00	7.23	15.14
				83.57	642.00	7.27	14.27
				90.00	307.00	8.38	20.60

Campanha							
2002				2006			
u	$ N(u) $	$\hat{\gamma}(u)$	D.p.	u	$ N(u) $	$\hat{\gamma}(u)$	D.p.
0.00	54.00	0.33	0.61	0.00	221.00	0.62	1.37
15.38	435.00	0.85	2.78	14.29	927.00	0.61	1.32
30.77	701.00	1.30	3.73	28.57	696.00	0.56	1.33
46.15	840.00	1.54	4.13	42.86	532.00	0.46	1.15
61.54	970.00	1.58	4.20	57.14	472.00	0.57	1.34
76.92	1099.00	1.50	3.89	71.43	431.00	0.55	1.30
92.31	1153.00	1.36	3.66	85.71	348.00	0.49	1.18
107.69	1064.00	1.52	3.72	100.00	275.00	0.40	0.90
123.08	1128.00	1.50	3.54	114.29	247.00	0.50	1.09
138.46	1011.00	1.25	3.08	128.57	206.00	0.55	1.12
153.85	787.00	1.43	3.14	142.86	155.00	0.44	0.74
169.23	601.00	1.52	2.92	157.14	49.00	0.64	0.96
184.62	438.00	1.86	3.16				
200.00	136.00	2.02	3.08				

Tabela 6: Estimativas dos parâmetros envolvidos nas estruturas da associação espacial dos modelos adotados para as concentrações de Cr nos musgos observadas em cada campanha.

Campanha	Modelo	Parâmetros					
		EM	MEEQ	SQR	$\hat{\tau}^2$	$\hat{\sigma}^2$	$\hat{\phi}$
1992	Exponencial	-0.048	3.38	5344.83	0.07	7.54	51.41
	Esférico	-0.005	1.85	4241.01	1.52	5.90	150.48
	Gaussiano	0.015	2.37	4374.57	2.23	5.17	70.79
1997	-	-	-	-	-	-	-
2002	Exponencial	0.003	1.67	214.69	0.07	1.43	15.47
	Esférico	0.007	1.91	190.33	0.28	1.21	47.01
	Gaussiano	0.010	2.25	192.97	0.40	1.09	22.16
2006	-	-	-	-	-	-	-

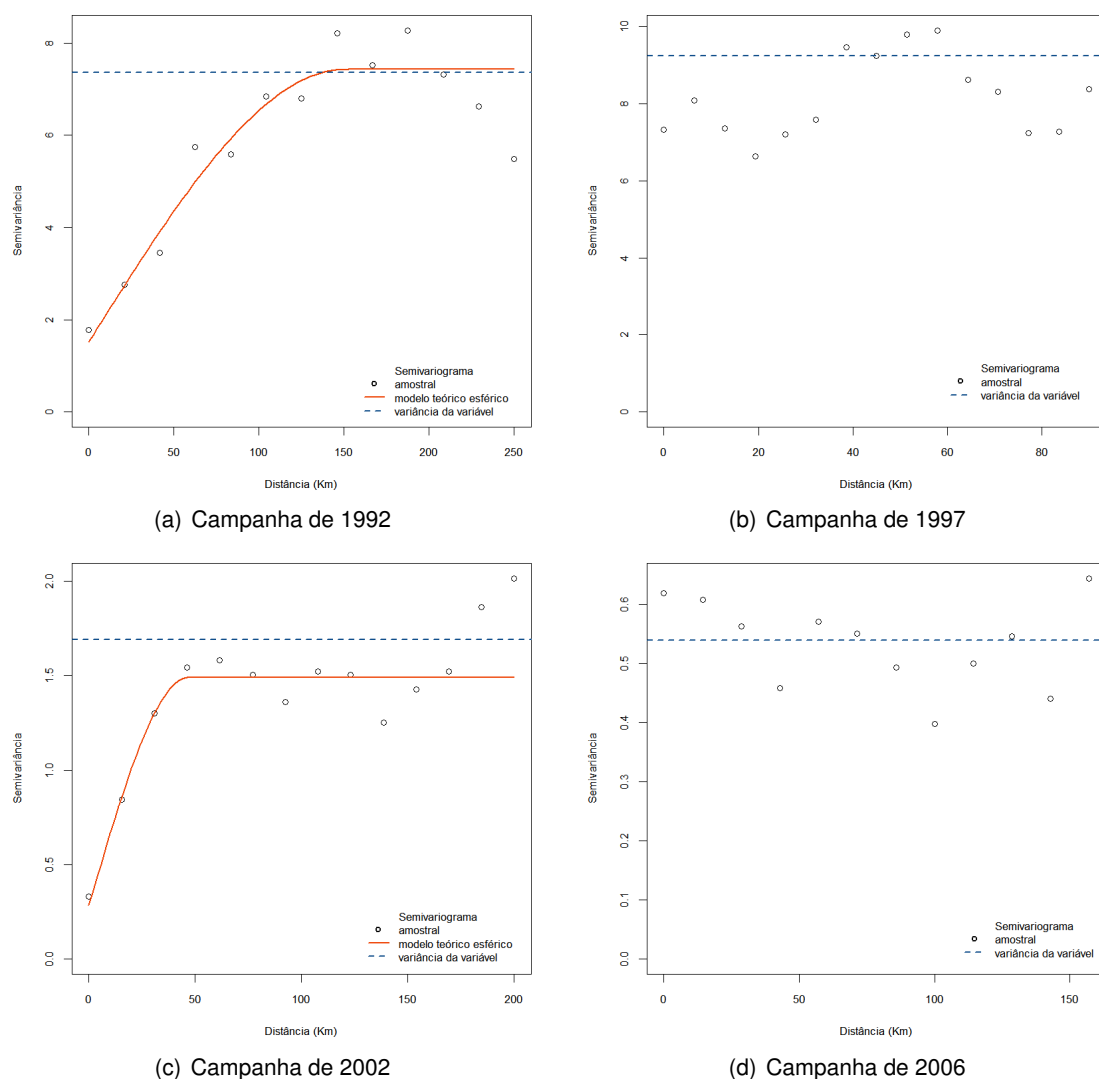


Figura 15: Variogramas amostrais das concentrações de Cr nos musgos observadas em cada campanha e respetivos modelos teóricos, estimados pelo método de mínimos quadrados.

Cobre

Tabela 7: Estimativas dos parâmetros envolvidos nas estruturas do valor médio e da anisotropia das concentrações de Cu nos musgos observadas em cada campanha.

Campanha	Tendência						Anisotropia	
	$\hat{\beta}_0$	D.p.	$\hat{\beta}_1$	D.p.	$\hat{\beta}_2$	D.p.	$\hat{\psi}_A$	$\hat{\psi}_R$
1992	8.33	0.34	-	-	-	-	0.07	2.47
1997	8.12	0.38	-	-	-	-	0.00	1.57
2002	6.75	0.23	-	-	-	-	0.37	1.46
2006	3.93	0.16	-	-	-	-	0.00	6.69

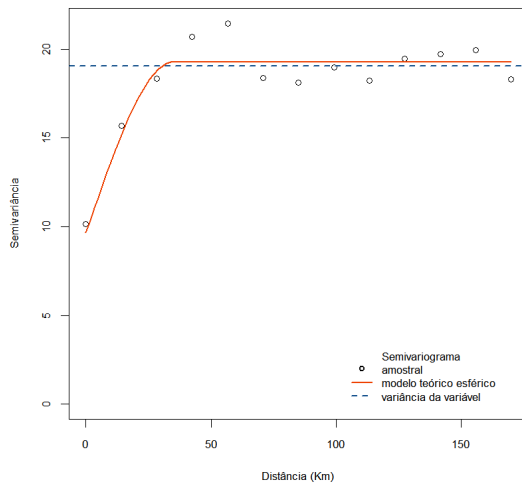
Tabela 8: Valores dos variogramas empíricos das concentrações de Cu nos musgos observadas em cada campanha, onde $|N(u)|$ representa o cardinal das diferentes classes de distâncias.

Campanha							
1992				1997			
u	$ N(u) $	$\hat{\gamma}(u)$	D.p.	u	$ N(u) $	$\hat{\gamma}(u)$	D.p.
0.00	93.00	10.16	13.79	0.00	40.00	24.35	41.58
14.17	557.00	15.69	29.49	14.17	417.00	26.52	46.30
28.33	810.00	18.35	33.36	28.33	594.00	21.50	42.57
42.50	862.00	20.70	35.11	42.50	675.00	20.52	41.89
56.67	1114.00	21.46	36.24	56.67	802.00	21.35	42.56
70.83	1166.00	18.37	32.81	70.83	886.00	20.90	41.73
85.00	1336.00	18.14	32.81	85.00	1068.00	22.20	47.19
99.17	1323.00	19.00	33.07	99.17	1064.00	23.83	47.02
113.33	1306.00	18.25	31.76	113.33	1035.00	25.48	52.37
127.50	1239.00	19.47	31.96	127.50	1085.00	24.65	49.35
141.67	1152.00	19.74	33.55	141.67	1046.00	26.71	48.68
155.83	987.00	19.95	34.62	155.83	1064.00	23.89	45.69
170.00	381.00	18.30	32.39	170.00	476.00	30.31	57.32

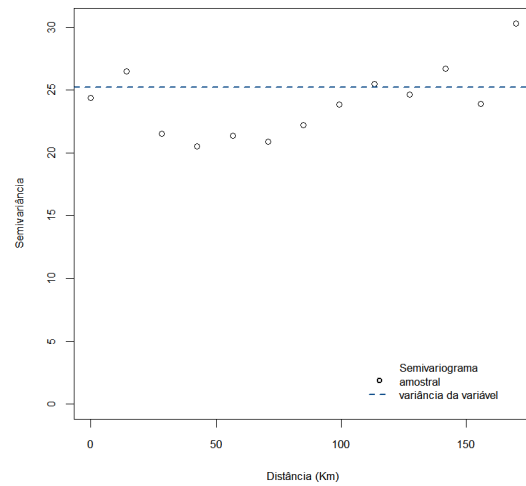
Campanha							
2002				2006			
u	$ N(u) $	$\hat{\gamma}(u)$	D.p.	u	$ N(u) $	$\hat{\gamma}(u)$	D.p.
0.00	26.00	2.66	3.55	0.00	124.00	2.31	3.76
16.67	309.00	5.21	9.67	12.31	657.00	2.47	3.69
33.33	475.00	7.18	12.84	24.62	655.00	2.60	4.00
50.00	579.00	7.01	13.86	36.92	528.00	2.50	3.81
66.67	683.00	8.04	14.80	49.23	415.00	2.43	3.66
83.33	773.00	7.60	15.03	61.54	429.00	2.10	3.21
100.00	800.00	6.57	12.54	73.85	316.00	2.24	3.15
133.33	884.00	8.51	14.55	86.15	310.00	2.38	3.55
150.00	856.00	8.86	16.09	98.46	246.00	2.42	3.84
166.67	767.00	9.12	16.21	110.77	218.00	2.91	4.40
183.33	649.00	8.27	15.40	123.08	201.00	2.57	4.32
200.00	587.00	6.99	12.87	135.38	148.00	2.77	4.66
216.67	491.00	7.35	13.59	147.69	118.00	1.57	2.46
233.33	472.00	8.69	14.89	160.00	34.00	1.41	2.02
250.00	214.00	8.19	15.62				

Tabela 9: Estimativas dos parâmetros envolvidos nas estruturas da associação espacial dos modelos adotados para as concentrações de Cu nos musgos observadas em cada campanha.

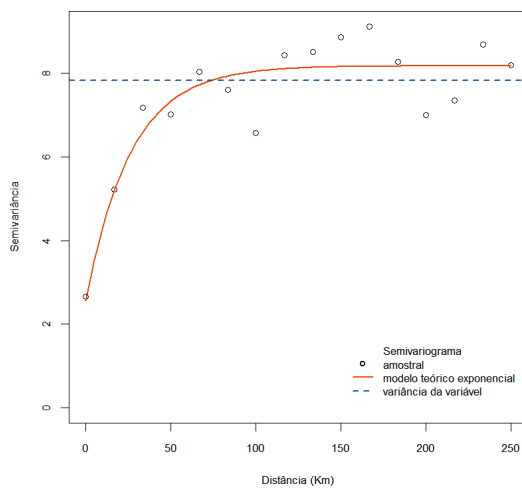
Campanha	Modelo	Parâmetros					
		EM	MEEQ	SQR	$\hat{\tau}^2$	$\hat{\sigma}^2$	$\hat{\phi}$
1992	Exponencial	-9.32e-15	0.98	19772.29	9.65	9.44	0.00
	Esférico	-0.008	1.01	12652.04	9.65	9.65	35.20
	Gaussiano	-0.019	1.07	12774.84	9.65	9.62	14.73
1997	-	-	-	-	-	-	-
2002	Exponencial	-0.004	0.88	4749.42	2.56	5.61	26.40
	Esférico	-0.017	0.75	5365.85	2.56	5.49	59.29
	Gaussiano	-7.36e-15	0.99	7972.85	2.56	5.31	0.00
2006	-	-	-	-	-	-	-



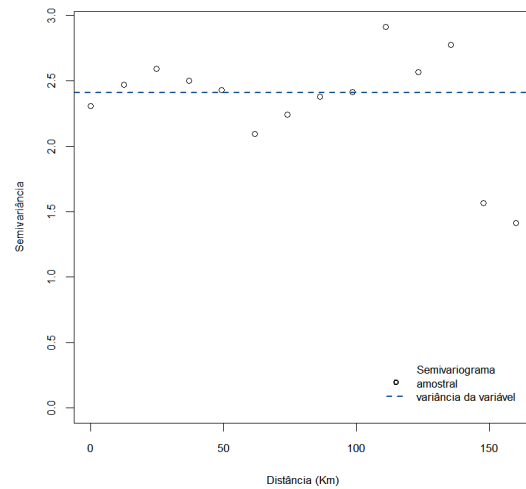
(a) Campanha de 1992



(b) Campanha de 1997



(c) Campanha de 2002



(d) Campanha de 2006

Figura 16: Variogramas amostrais das concentrações de Cu nos musgos observadas em cada campanha e respetivos modelos teóricos, estimados pelo método de mínimos quadrados.

Ferro

Tabela 10: Estimativas dos parâmetros envolvidos nas estruturas do valor médio e da anisotropia das concentrações de Fe nos musgos observadas em cada campanha.

Campanha	Tendência						Anisotropia	
	$\hat{\beta}_0$	D.p.	$\hat{\beta}_1$	D.p.	$\hat{\beta}_2$	D.p.	$\hat{\psi}_A$	$\hat{\psi}_R$
1992	1266.48	77.75	9.24	1.37	-1.99	0.54	0.00	1.43
1997	1964.36	118.89	8.60	2.10	-6.21	0.84	0.00	1.67
2002	772.48	47.92	0.90	0.84	-1.60	0.34	9.33e-09	1.00
2006	295.58	60.98	1.51	0.64	1.29	0.66	0.47	6.47

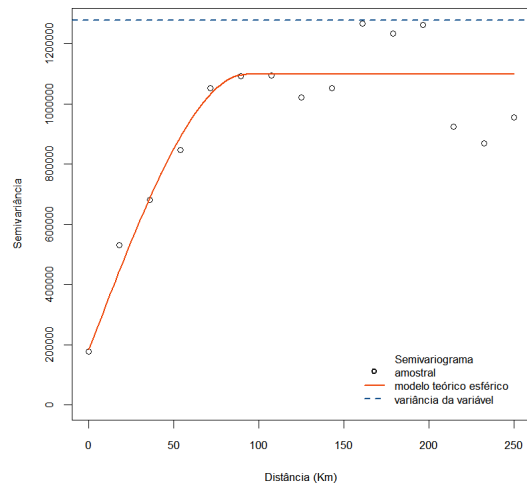
Tabela 11: Valores dos variogramas empíricos das concentrações de Fe nos musgos observadas em cada campanha, onde $|N(u)|$ representa o cardinal das diferentes classes de distâncias.

Campanha							
1992				1997			
u	$ N(u) $	$\hat{\gamma}(u)$	D.p.	u	$ N(u) $	$\hat{\gamma}(u)$	D.p.
0.00	77.00	177477.72	301527.14	0.00	44.00	764576.78	1396801.45
17.86	562.00	529572.37	1904530.18	13.64	439.00	1460420.04	2932262.83
35.71	759.00	681291.92	2047103.31	27.27	618.00	1404205.53	2885965.23
53.57	879.00	846094.27	2698723.42	40.91	672.00	1830426.64	4023317.21
71.43	1033.00	1052999.73	3107841.50	54.55	821.00	2103626.64	4372795.19
89.29	1198.00	1090984.15	3129439.79	68.18	921.00	2172733.22	4440024.53
107.14	1211.00	1093871.60	2837570.05	81.82	1030.00	2173696.54	4436993.47
125.00	1226.00	1021719.20	2461887.64	95.45	1090.00	2275886.62	4346369.23
142.86	1239.00	1052050.09	2493124.74	109.09	1057.00	2429239.17	4485602.77
160.71	1155.00	1266777.99	3021366.24	122.73	1057.00	2576516.51	4515783.03
178.57	1063.00	1234366.64	3070536.12	136.36	1090.00	2178828.59	3773989.48
196.43	848.00	1263076.54	2970047.32	150.00	1106.00	2274197.39	3923425.03
214.29	738.00	923412.54	2172300.05	163.64	970.00	2266283.58	3569484.87
232.14	661.00	867879.50	2032143.14	177.27	799.00	2591693.23	4263163.75
250.00	304.00	955514.22	2237630.58	190.91	729.00	2813162.67	4468078.09
				204.55	632.00	3023670.53	4300323.77
				218.18	544.00	3047660.71	4392912.42
				231.82	406.00	2763590.85	3900695.54
				245.45	353.00	2609171.42	3886679.77
				259.09	283.00	2765938.36	4390016.20
				272.73	229.00	3237315.39	4375154.73
				286.36	161.00	4082590.66	4841081.34
				300.00	56.00	1897758.27	2662380.41

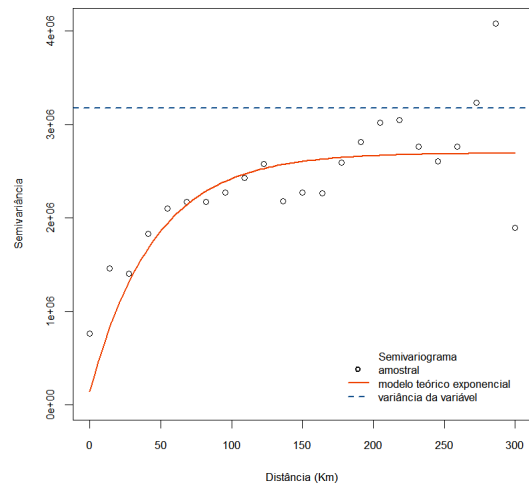
Campanha							
2002				2006			
u	$ N(u) $	$\hat{\gamma}(u)$	D.p.	u	$ N(u) $	$\hat{\gamma}(u)$	D.p.
0.00	34.00	128960.72	276431.11	0.00	30.00	85378.17	179855.03
22.22	384.00	206463.24	566214.23	20.00	405.00	97682.91	168714.64
44.44	552.00	287218.27	750032.23	40.00	583.00	68590.82	135561.77
66.67	670.00	279530.46	723477.60	60.00	725.00	90602.35	163250.31
88.89	776.00	273992.57	710445.74	80.00	734.00	84431.68	157262.34
111.11	799.00	310655.73	775832.02	100.00	675.00	100454.18	173280.27
133.33	874.00	312984.15	790751.61	120.00	569.00	80869.72	144926.27
155.56	896.00	372583.12	864682.32	140.00	463.00	107193.49	175443.25
177.78	750.00	365249.72	863794.27	160.00	260.00	104641.81	185572.46
200.00	682.00	314543.24	729237.42	180.00	70.00	70476.01	126832.51
222.22	598.00	307470.32	746037.73				
244.44	537.00	300880.35	747560.94				
266.67	479.00	346268.08	886360.64				
288.89	453.00	364678.47	875656.16				
311.11	431.00	302896.45	762928.63				
333.33	411.00	264137.19	651960.96				
355.56	343.00	335076.14	840390.98				
377.78	318.00	396176.49	925907.40				
400.00	121.00	384008.33	930153.81				

Tabela 12: Estimativas dos parâmetros envolvidos nas estruturas da associação espacial dos modelos adotados para as concentrações de Fe nos musgos observadas em cada campanha.

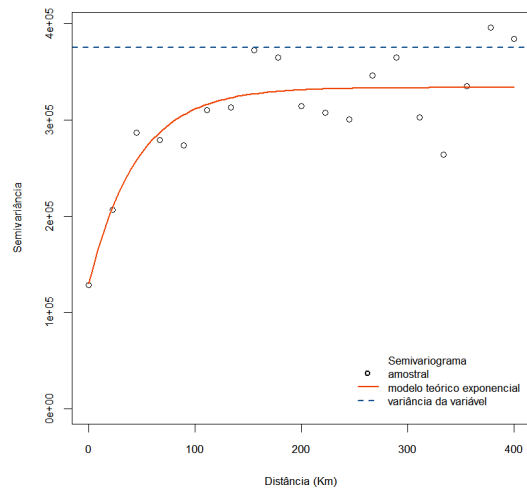
Campanha	Modelo	Parâmetros					
		EM	MEEQ	SQR	$\hat{\tau}^2$	$\hat{\sigma}^2$	$\hat{\phi}$
1992	Exponencial	2.03	2.21	1659e+11	183226.01	939988.04	37.56
	Esférico	-0.12	1.76	1554e+11	195346.01	914531.26	92.61
	Gaussiano	7.48	4.05	1756e+11	197548.01	912145.40	41.28
1997	Exponencial	-2.55	1.01	1294e+12	144600.77	2551297.02	44.78
	Esférico	6.07	2.39	8820e+12	1620627.89	3343313.28	876.20
	Gaussiano	3.08	1.01	2874e+12	764578.76	1597998.51	0.00
2002	Exponencial	0.61	1.04	9758e+9	129960.70	203747.01	45.12
	Esférico	7.90	∞	1882e+10	315508.29	0.00	50.20
	Gaussiano	-1.03	1.31	1763e+10	128960.74	187177.15	0.00
2006	-	-	-	-	-	-	-



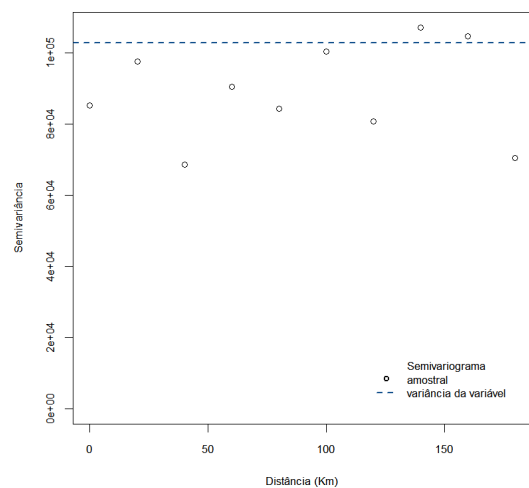
(a) Campanha de 1992



(b) Campanha de 1997



(c) Campanha de 2002



(d) Campanha de 2006

Figura 17: Variogramas amostrais das concentrações de Fe nos musgos observadas em cada campanha e respetivos modelos teóricos, estimados pelo método de mínimos quadrados.

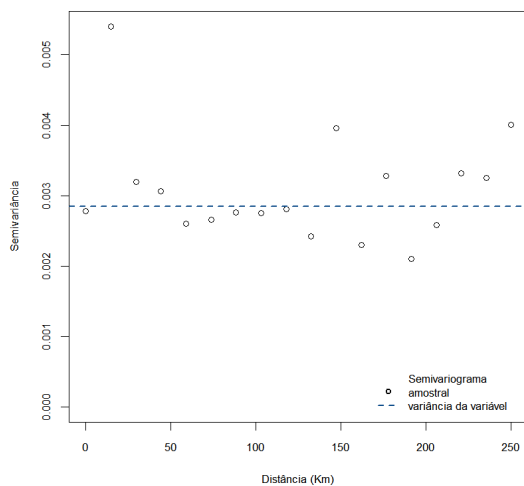
Mercúrio

Tabela 13: Estimativas dos parâmetros envolvidos nas estruturas do valor médio e da anisotropia das concentrações de Hg nos musgos observadas em cada campanha.

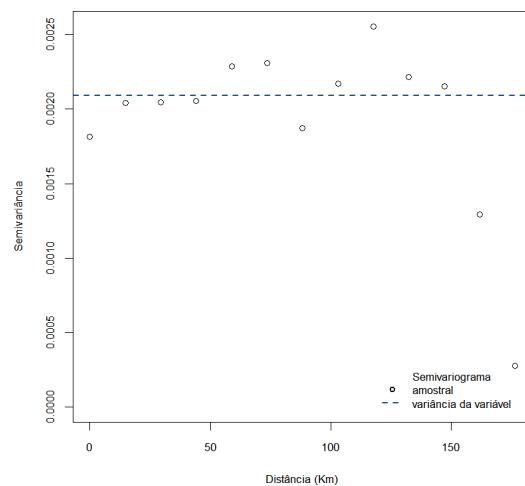
Campanha	Tendência						Anisotropia	
	$\hat{\beta}_0$	D.p.	$\hat{\beta}_1$	D.p.	$\hat{\beta}_2$	D.p.	$\hat{\psi}_A$	$\hat{\psi}_R$
2002	5.13e-02	3.16e-03	-1.57e-04	5.48e-05	2.67e-05	2.18e-05	1.39	7.54
2006	0.06	0.01	-	-	-	-	0.73	34.81

Tabela 14: Valores dos variogramas empíricos das concentrações de Hg nos musgos observadas em cada campanha, onde $|N(u)|$ representa o cardinal das diferentes classes de distâncias.

Campanha							
2002				2006			
u	$ N(u) $	$\hat{\gamma}(u)$	D.p.	u	$ N(u) $	$\hat{\gamma}(u)$	D.p.
0.00	109.00	0.0013	0.0024	0.00	399.00	0.0018	0.0035
14.71	613.00	0.0011	0.0021	14.71	896.00	0.0020	0.0039
29.41	662.00	0.0015	0.0025	29.41	789.00	0.0020	0.0040
44.12	561.00	0.0016	0.0029	44.12	707.00	0.0021	0.0039
58.82	528.00	0.0014	0.0026	58.82	589.00	0.0023	0.0041
73.53	545.00	0.0017	0.0032	73.53	460.00	0.0023	0.0041
88.24	507.00	0.0017	0.0031	88.24	323.00	0.0019	0.0034
102.94	485.00	0.0017	0.0028	102.94	205.00	0.0022	0.0038
117.65	460.00	0.0016	0.0028	117.65	123.00	0.0026	0.0047
132.35	448.00	0.0015	0.0025	132.35	93.00	0.0022	0.0037
147.06	424.00	0.0016	0.0031	147.06	50.00	0.0022	0.0033
161.76	421.00	0.0015	0.0030	161.76	18.00	0.0013	0.0030
176.47	422.00	0.0015	0.0024	176.47	4.00	0.0003	0.0002
191.18	384.00	0.0013	0.0022				
205.88	365.00	0.0015	0.0025				
220.59	347.00	0.0016	0.0028				
235.29	325.00	0.0016	0.0030				
250.00	158.00	0.0018	0.0029				



(a) Campanha de 2002



(b) Campanha de 2006

Figura 18: Variogramas amostrais das concentrações de Hg nos musgos observadas em cada campanha.

Mangânês

Tabela 15: Estimativas dos parâmetros envolvidos nas estruturas do valor médio e da anisotropia das concentrações de Mn nos musgos observadas em cada campanha.

Campanha	Tendência						Anisotropia	
	$\hat{\beta}_0$	D.p.	$\hat{\beta}_1$	D.p.	$\hat{\beta}_2$	D.p.	$\hat{\psi}_A$	$\hat{\psi}_R$
1992	170.55	9.75	1.04	0.17	-0.30	0.07	0.00	1.24
1997	187.35	10.21	0.66	0.18	-0.14	0.07	0.00	1.00
2002	148.95	7.14	-	-	-	-	0.00	1.00
2006	74.85	19.33	0.61	0.20	0.48	0.21	2.04e-06	1.00

Tabela 16: Valores dos variogramas empíricos das concentrações de Mn nos musgos observadas em cada campanha, onde $|N(u)|$ representa o cardinal das diferentes classes de distâncias.

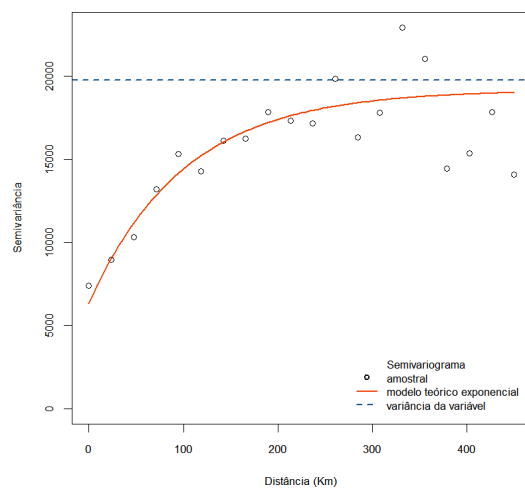
Campanha							
1992				1997			
u	$ N(u) $	$\hat{\gamma}(u)$	D.p.	u	$ N(u) $	$\hat{\gamma}(u)$	D.p.
0.00	123.00	7385.24	14304.04	0.00	63.00	15784.27	32921.79
23.68	752.00	8958.97	24574.44	20.59	521.00	15126.81	29978.36
47.37	994.00	10320.58	27292.80	41.18	710.00	17294.31	32870.81
71.05	1199.00	13201.98	32228.68	61.76	833.00	18710.64	34344.05
94.74	1417.00	15299.44	34800.52	82.35	974.00	18289.07	31811.48
118.42	1466.00	14284.04	32558.94	102.94	1102.00	16381.16	30696.43
142.11	1503.00	16129.45	35814.23	123.53	1203.00	14225.72	28211.21
165.79	1341.00	16245.33	36420.48	144.12	1132.00	15129.94	29508.78
189.47	1201.00	17836.59	41180.92	164.71	1062.00	18250.06	31795.71
213.16	1016.00	17295.32	37588.04	185.29	956.00	18212.76	33934.41
236.84	828.00	17168.72	36480.31	205.88	845.00	15578.86	29341.23
260.53	758.00	19826.17	41909.13	226.47	769.00	13109.78	27907.21
284.21	654.00	16326.84	32755.46	247.06	780.00	17464.99	32535.43
307.89	500.00	17780.35	33828.90	267.65	647.00	22555.59	38552.95
331.58	391.00	22913.43	38445.61	288.24	565.00	23679.37	36666.78
355.26	265.00	21029.40	32430.77	308.82	563.00	19744.55	33090.89
378.95	158.00	14418.87	22495.92	329.41	524.00	16718.12	28650.05
402.63	88.00	15350.34	24556.42	350.00	254.00	18157.64	28309.00
426.32	46.00	17826.37	23621.12				
450.00	5.00	14084.93	9142.66				

Campanha							
2002				2006			
u	$ N(u) $	$\hat{\gamma}(u)$	D.p.	u	$ N(u) $	$\hat{\gamma}(u)$	D.p.
0.00	28.00	3812.24	4444.69	0.00	29.00	5261.87	11221.50
20.83	334.00	4818.45	7991.82	20.00	411.00	8644.26	16277.31
41.67	518.00	6048.45	9125.48	40.00	591.00	9217.02	17027.91

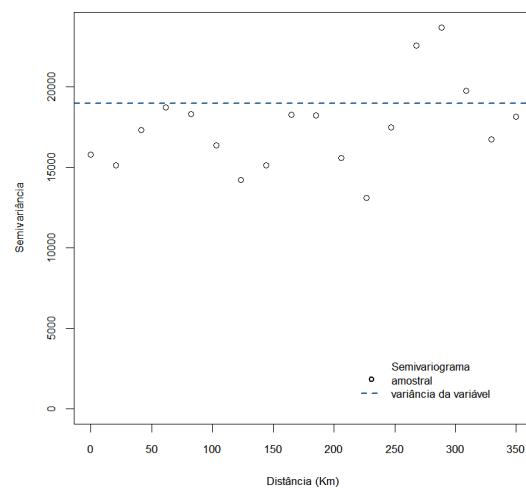
Campanha							
2002				2006			
u	$ N(u) $	$\hat{\gamma}(u)$	D.p.	u	$ N(u) $	$\hat{\gamma}(u)$	D.p.
62.50	607.00	7460.55	9896.14	60.00	728.00	8787.44	15831.89
83.33	710.00	8873.81	11243.24	80.00	735.00	8559.30	14410.60
104.17	775.00	8241.49	11575.38	100.00	672.00	8510.58	14216.57
125.00	846.00	7884.18	11065.75	120.00	568.00	8644.76	15384.83
145.83	825.00	7970.29	11595.21	140.00	456.00	8948.05	14041.70
166.67	794.00	7928.89	11606.87	160.00	255.00	10837.12	14170.80
187.50	710.00	9529.35	12615.38	180.00	109.00	10412.43	10816.58
208.33	648.00	8591.87	11980.51	200.00	57.00	11697.36	11191.24
229.17	562.00	7309.02	11023.05	220.00	14.00	14120.62	12527.38
250.00	254.00	8366.81	12316.09				

Tabela 17: Estimativas dos parâmetros envolvidos nas estruturas da associação espacial dos modelos adotados para as concentrações de Mn nos musgos observadas em cada campanha.

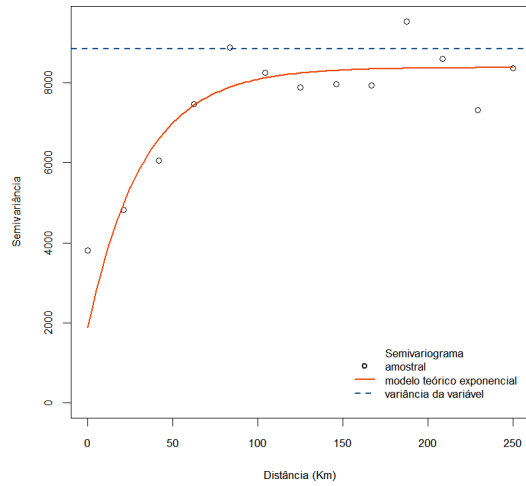
Campanha	Modelo	Parâmetros					
		EM	MEEQ	SQR	$\hat{\tau}^2$	$\hat{\sigma}^2$	$\hat{\phi}$
1992	Exponencial	0.13	1.47	2331e+7	6330.96	12814.36	100.41
	Esférico	0.16	1.62	2589e+7	8045.07	10139.40	236.75
	Gaussiano	0.24	1.50	2738e+7	9256.33	8799.11	110.23
1997	-	-	-	-	-	-	-
2002	Exponencial	0.92	1.14	2947e+6	1875.27	6502.12	32.25
	Esférico	2.53e-13	1.12	8276e+6	3812.24	4108.88	0.00
	Gaussiano	2.34e-13	1.12	8276e+6	3812.23	4108.89	0.00
2006	Exponencial	0.02	1.05	1303e+6	5236.40	3748.91	7.84
	Esférico	-0.08	1.09	1981e+6	5134.96	3574.42	11.80
	Gaussiano	-0.11	1.10	1576e+6	3380.61	3194.07	5.62



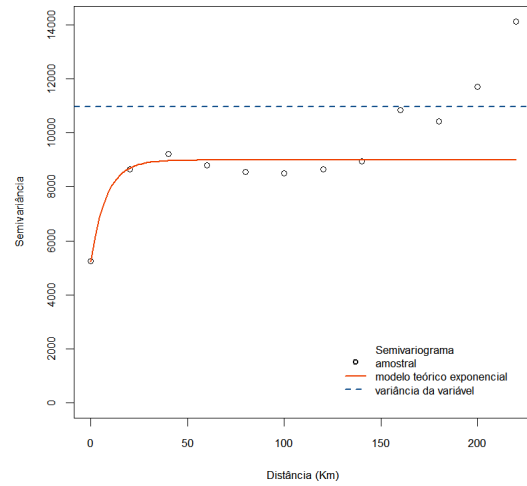
(a) Campanha de 1992



(b) Campanha de 1997



(c) Campanha de 2002



(d) Campanha de 2006

Figura 19: Variogramas amostrais das concentrações de Mn nos musgos observadas em cada campanha e respectivos modelos teóricos, estimados pelo método de mínimos quadrados.

Níquel

Tabela 18: Estimativas dos parâmetros envolvidos nas estruturas do valor médio e da anisotropia das concentrações de Ni nos musgos observadas em cada campanha.

Campanha	Tendência						Anisotropia	
	$\hat{\beta}_0$	D.p.	$\hat{\beta}_1$	D.p.	$\hat{\beta}_2$	D.p.	$\hat{\psi}_A$	$\hat{\psi}_R$
1992	2.21	0.13	0.005	0.002	-0.003	0.001	0.00	1.09
1997	10.95	0.52	-	-	-	-	1.23	5.16
2002	1.58	0.12	0.001	0.002	-0.002	0.001	6.38e-09	1.00

Tabela 19: Valores dos variogramas empíricos das concentrações de Ni nos musgos observadas em cada campanha, onde $|N(u)|$ representa o cardinal das diferentes classes de distâncias.

Campanha							
1992				1997			
u	$ N(u) $	$\hat{\gamma}(u)$	D.p.	u	$ N(u) $	$\hat{\gamma}(u)$	D.p.
0.00	58.00	0.75	0.96	0.00	344.00	47.44	66.18
18.75	457.00	1.29	2.65	23.33	1513.00	41.20	62.06
37.50	625.00	1.99	4.24	46.67	1462.00	39.31	61.47
56.25	735.00	2.55	5.27	70.00	1248.00	43.17	68.89
75.00	852.00	2.36	4.33	93.33	1176.00	42.31	67.67
93.75	958.00	2.34	4.74	116.67	1070.00	41.85	67.84
112.50	1062.00	2.61	5.05	140.00	1077.00	46.87	73.11
131.25	1029.00	2.75	4.68	163.33	1007.00	48.77	75.68
150.00	1037.00	2.93	5.00	186.67	959.00	46.40	73.06
168.75	937.00	2.81	4.9	210.00	895.00	49.66	78.52
187.50	816.00	3.03	5.15	233.33	754.00	48.93	79.59
206.25	757.00	2.72	4.69	256.67	623.00	48.70	78.64
225.00	729.00	3.27	5.95	280.00	273.00	49.61	79.99

Campanha							
1992				1997			
u	$ N(u) $	$\hat{\gamma}(u)$	D.p.	u	$ N(u) $	$\hat{\gamma}(u)$	D.p.
243.75	641.00	3.29	5.75				
262.50	534.00	3.41	6.18				
281.25	542.00	3.11	5.36				
300.00	241.00	2.41	4.35				

Campanha							
2002							
u	$ N(u) $	$\hat{\gamma}(u)$	D.p.				
0.00	27.00	1.34	2.96				
20.83	318.00	1.61	3.06				
41.67	497.00	2.08	4.10				
62.50	589.00	1.95	3.82				
83.33	681.00	1.94	3.89				
104.17	743.00	2.27	4.29				
125.00	810.00	2.15	4.19				
145.83	769.00	1.95	3.60				
166.67	753.00	2.41	4.30				
187.50	675.00	2.26	4.23				
208.33	606.00	2.48	4.60				
229.17	540.00	1.95	3.93				
250.00	233.00	1.80	3.67				

Tabela 20: Estimativas dos parâmetros envolvidos nas estruturas da associação espacial dos modelos adotados para as concentrações de Ni nos musgos observadas em cada campanha.

Campanha	Modelo	Parâmetros					
		EM	MEEQ	SQR	$\hat{\tau}^2$	$\hat{\sigma}^2$	$\hat{\phi}$
1992	Exponencial	0.008	1.48	549.44	0.75	2.34	61.38
	Esférico	0.006	1.19	626.31	1.56	1.55	232.60
	Gaussiano	0.006	1.19	707.78	1.74	1.33	104.99
1997	-	-	-	-	-	-	-
2002	Exponencial	-0.002	1.08	247.70	1.34	0.86	38.94
	Esférico	-0.001	1.08	268.11	1.11	1.04	55.97
	Gaussiano	-0.002	1.10	264.64	1.30	0.85	29.95

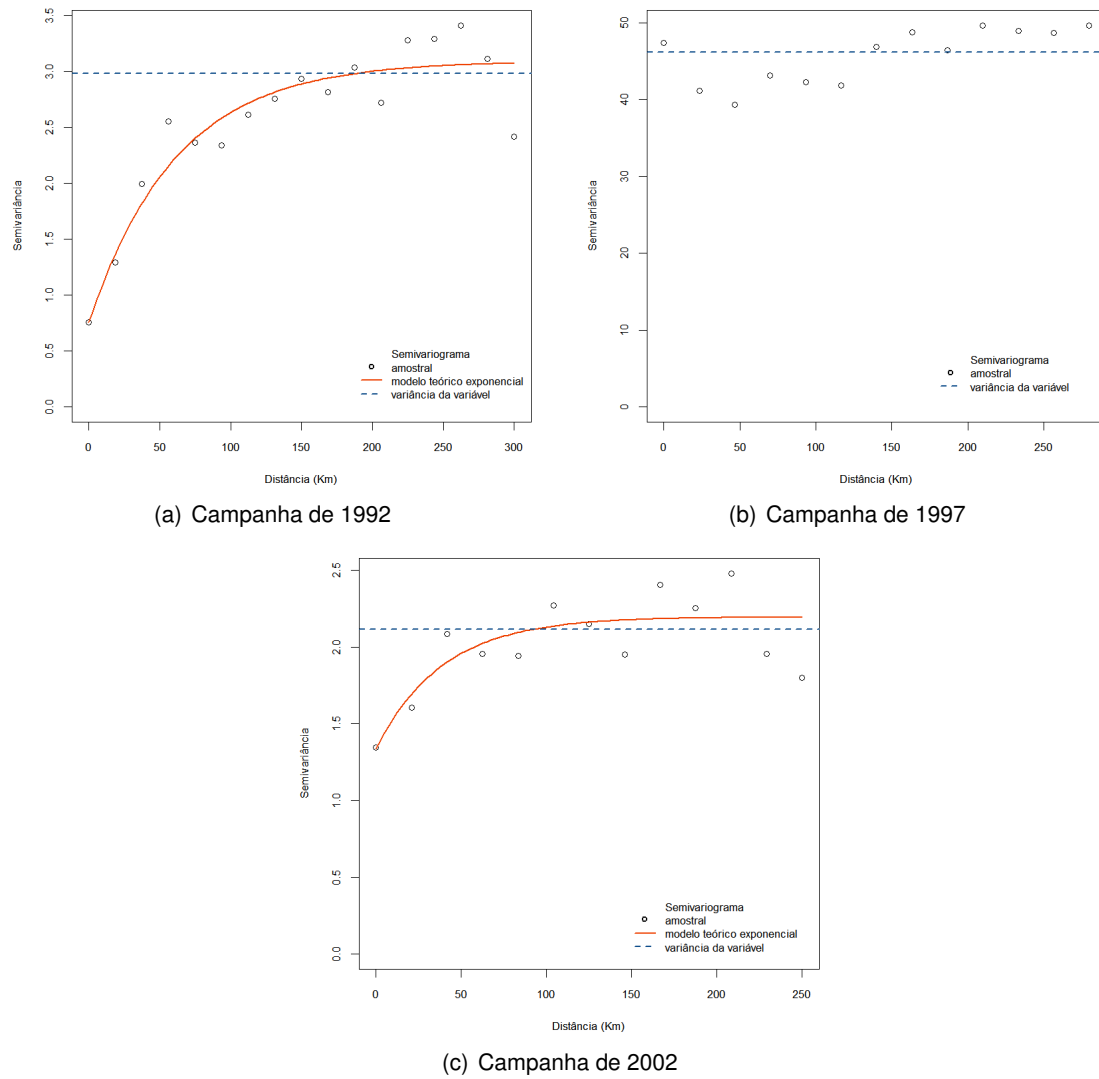


Figura 20: Variogramas amostrais das concentrações de Ni nos musgos observadas em cada campanha e respetivos modelos teóricos, estimados pelo método de mínimos quadrados.

Chumbo

Tabela 21: Estimativas dos parâmetros envolvidos nas estruturas do valor médio e da anisotropia das concentrações de Pb nos musgos observadas em cada campanha.

Campanha	Tendência						Anisotropia	
	$\hat{\beta}_0$	D.p.	$\hat{\beta}_1$	D.p.	$\hat{\beta}_2$	D.p.	$\hat{\psi}_A$	$\hat{\psi}_R$
1992	15.00	0.70	-0.05	0.01	0.01	0.00	0.59	2.63
1997	19.69	1.13	-	-	-	-	0.00	1.60
2002	5.36	0.85	-	-	-	-	0.00	1.00
2006	1.04	0.32	-	-	-	-	0.00	1.00

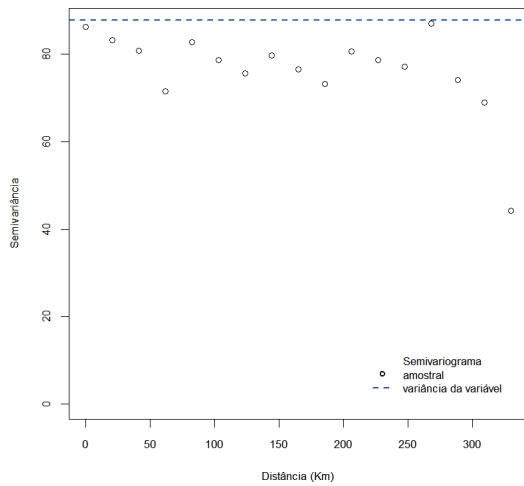
Tabela 22: Valores dos variogramas empíricos das concentrações de Pb nos musgos observadas em cada campanha, onde $|N(u)|$ representa o cardinal das diferentes classes de distâncias.

Campanha							
1992				1997			
u	$ N(u) $	$\hat{\gamma}(u)$	D.p.	u	$ N(u) $	$\hat{\gamma}(u)$	D.p.
0.00	182.00	86.32	131.66	0.00	33.00	104.55	116.10
20.62	914.00	83.30	132.22	12.50	354.00	158.96	355.04
41.25	1321.00	80.75	120.09	25.00	506.00	217.24	595.38
61.88	1519.00	71.61	110.77	37.50	587.00	258.80	647.41
82.50	1688.00	82.80	129.84	50.00	676.00	253.35	628.16
103.12	1750.00	78.69	124.53	62.50	788.00	283.69	696.38
123.75	1564.00	75.66	115.93	75.00	825.00	215.73	522.71
144.38	1186.00	79.80	114.82	87.50	953.00	213.74	543.00
165.00	978.00	76.57	124.45	100.00	951.00	232.39	563.34
185.62	787.00	73.26	126.76	112.50	953.00	238.22	599.87
206.25	611.00	80.58	152.95	125.00	934.00	232.21	578.18
226.88	482.00	78.73	133.62	137.50	965.00	227.95	587.71
247.50	339.00	77.22	139.13	150.00	985.00	213.35	542.70
268.12	280.00	87.05	161.31	162.50	901.00	199.64	473.81
288.75	178.00	74.18	125.42	175.00	742.00	212.12	512.00
309.38	124.00	68.95	112.94	187.50	640.00	230.07	534.64
330.00	37.00	44.28	65.91	200.00	321.00	195.91	363.87

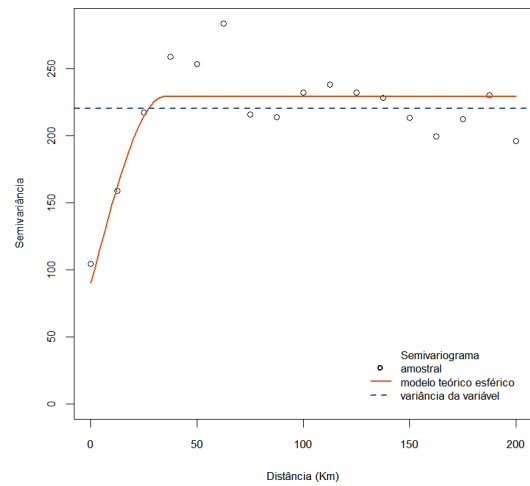
Campanha							
2002				2006			
u	$ N(u) $	$\hat{\gamma}(u)$	D.p.	u	$ N(u) $	$\hat{\gamma}(u)$	D.p.
0.00	29.00	20.82	29.71	0.00	32.00	3.27	6.23
20.83	338.00	65.53	452.22	20.00	410.00	2.81	5.25
41.67	521.00	77.21	452.59	40.00	584.00	2.67	4.83
62.50	615.00	92.09	560.80	60.00	719.00	2.49	4.75
83.33	721.00	82.03	562.24	80.00	735.00	2.65	4.97
104.17	786.00	108.18	676.99	100.00	674.00	2.82	5.17
125.00	859.00	80.25	557.09	120.00	569.00	2.88	5.06
145.83	832.00	159.59	839.22	140.00	466.00	2.55	5.14
166.67	799.00	113.78	680.53	160.00	258.00	2.96	5.85
187.50	718.00	150.54	801.50	180.00	69.00	2.26	4.81
208.33	652.00	82.91	514.46				
229.17	566.00	66.25	499.14				
250.00	255.00	97.59	644.08				

Tabela 23: Estimativas dos parâmetros envolvidos nas estruturas da associação espacial dos modelos adotados para as concentrações de Pb nos musgos observadas em cada campanha.

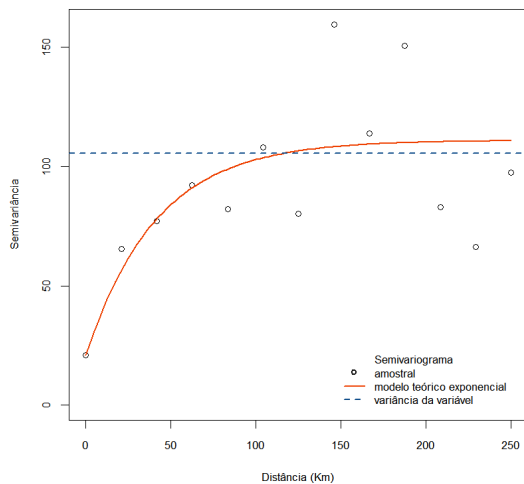
Campanha	Modelo	Parâmetros					
		EM	MEEQ	SQR	$\hat{\tau}^2$	$\hat{\sigma}^2$	$\hat{\phi}$
1992	-	-	-	-	-	-	-
	Exponencial	-4.89e-14	0.97	7064e+3	104.55	122.01	0.00
1997	Esférico	0.01	1.13	5343e+3	89.97	139.17	34.40
	Gaussiano	0.04	1.16	5347e+3	97.69	131.34	15.57
	Exponencial	0.07	1.66	5879e+3	20.80	90.19	41.38
2002	Esférico	0.10	2.24	6069e+3	20.80	88.85	116.27
	Gaussiano	9.11e-15	1.04	7081e+3	20.80	81.04	0.00
2006	-	-	-	-	-	-	-



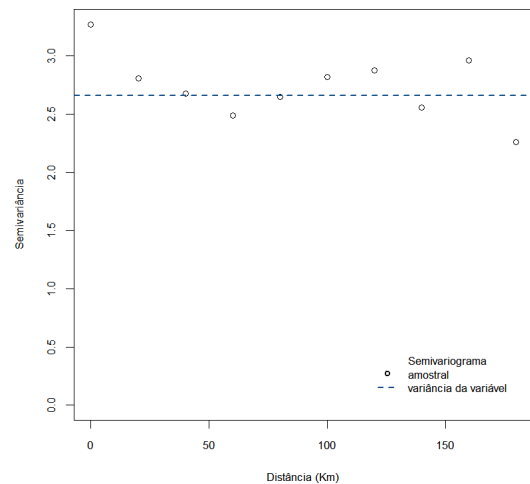
(a) Campanha de 1992



(b) Campanha de 1997



(c) Campanha de 2002



(d) Campanha de 2006

Figura 21: Variogramas amostrais das concentrações de Pb nos musgos observadas em cada campanha e respetivos modelos teóricos, estimados pelo método de mínimos quadrados.

Zinco

Tabela 24: Estimativas dos parâmetros envolvidos nas estruturas do valor médio e da anisotropia das concentrações de Zn nos musgos observadas em cada campanha.

Campanha	Tendência						Anisotropia	
	$\hat{\beta}_0$	D.p.	$\hat{\beta}_1$	D.p.	$\hat{\beta}_2$	D.p.	$\hat{\psi}_A$	$\hat{\psi}_R$
1992	34.13	1.54	-0.10	0.03	0.02	0.01	2.42e-10	1.00
1997	45.24	1.94	-0.15	0.03	0.03	0.01	0.79	2.69
2002	32.02	1.14	-	-	-	-	0.14	2.43
2006	15.33	2.26	-	-	-	-	0.00	1.00

Tabela 25: Valores dos variogramas empíricos das concentrações de Zn nos musgos observadas em cada campanha, onde $|N(u)|$ representa o cardinal das diferentes classes de distâncias.

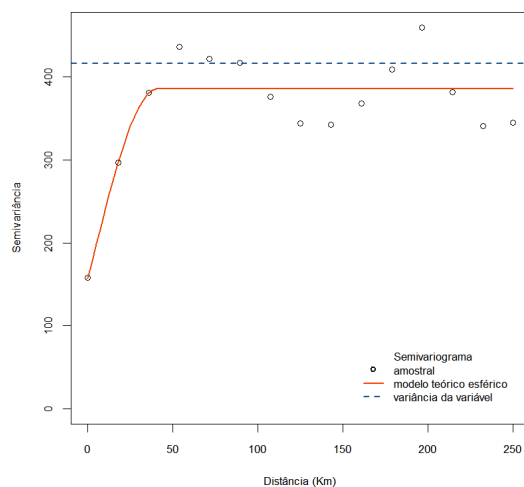
Campanha							
1992				1997			
u	$ N(u) $	$\hat{\gamma}(u)$	D.p.	u	$ N(u) $	$\hat{\gamma}(u)$	D.p.
0.00	44.00	157.81	218.07	0.00	82.00	547.15	1090.17
17.86	383.00	296.70	494.51	14.29	546.00	724.62	1491.04
35.71	553.00	380.74	703.50	28.57	770.00	660.27	1466.01
53.57	652.00	435.71	847.44	42.86	1033.00	630.11	1244.10
71.43	727.00	421.29	805.57	57.14	1054.00	589.46	1245.28
89.29	849.00	416.96	798.95	71.43	1119.00	575.63	1089.88
107.14	884.00	375.83	759.92	85.71	1078.00	612.67	1148.81
125.00	1003.00	343.53	644.08	100.00	1052.00	565.25	1155.49
142.86	936.00	341.93	695.93	114.29	1011.00	617.27	1274.33
160.71	899.00	367.46	696.06	128.57	901.00	669.38	1340.52
178.57	834.00	408.59	743.32	142.86	829.00	688.08	1376.41
196.43	685.00	459.31	874.12	157.14	814.00	670.55	1373.75
214.29	682.00	381.25	770.43	171.43	695.00	626.76	1340.52
232.14	655.00	340.76	636.74	185.71	592.00	462.13	1033.93
250.00	307.00	344.68	616.57	200.00	257.00	522.69	1197.18

Campanha							
2002				2006			
u	$ N(u) $	$\hat{\gamma}(u)$	D.p.	u	$ N(u) $	$\hat{\gamma}(u)$	D.p.
0.00	53.00	125.55	181.86	6.67	29.00	181.22	269.90
16.67	452.00	184.96	328.81	13.33	90.00	108.26	153.68
33.33	686.00	213.34	360.88	20.00	121.00	140.81	236.92
50.00	889.00	216.85	375.97	26.67	161.00	129.63	221.49
66.67	1001.00	226.79	381.83	33.33	169.00	148.32	277.60
83.33	1124.00	216.84	355.69	40.00	155.00	123.10	252.04
100.00	1173.00	211.35	348.69	46.67	218.00	137.80	240.42
116.67	1128.00	219.80	358.04	53.33	186.00	150.37	275.53

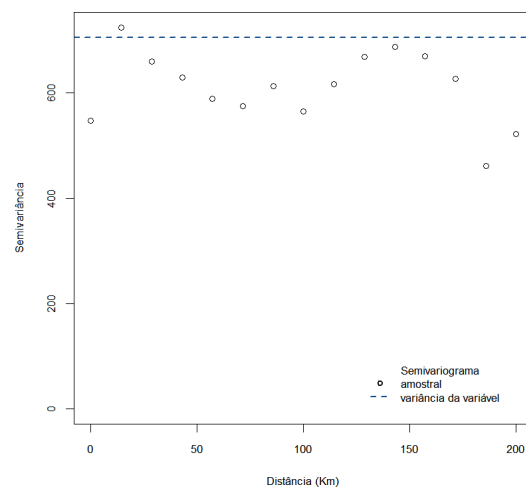
Campanha							
2002				2006			
u	$ N(u) $	$\hat{\gamma}(u)$	D.p.	u	$ N(u) $	$\hat{\gamma}(u)$	D.p.
133.33	1118.00	266.10	432.09	60.00	222.00	127.73	234.60
150.00	1025.00	238.32	360.10	66.67	241.00	166.87	303.41
166.67	727.00	254.38	401.14	73.33	224.00	146.42	265.51
183.33	529.00	252.98	400.94	80.00	220.00	152.13	245.92
200.00	211.00	300.65	450.35	86.67	224.00	131.03	255.43
				93.33	217.00	136.79	252.02
				100.00	213.00	161.07	269.60
				106.67	185.00	186.94	313.59
				113.33	195.00	165.83	319.32
				120.00	85.00	214.49	344.77

Tabela 26: Estimativas dos parâmetros envolvidos nas estruturas da associação espacial dos modelos adotados para as concentrações de Zn nos musgos observadas em cada campanha.

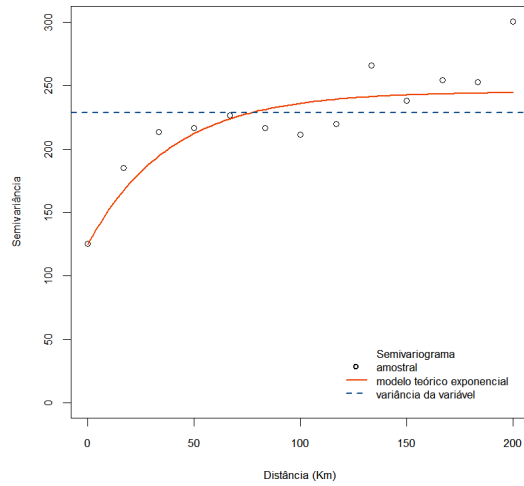
Campanha	Modelo	Parâmetros					
		EM	MEEQ	SQR	$\hat{\tau}^2$	$\hat{\sigma}^2$	$\hat{\phi}$
1992	Exponencial	0.007	1.02	1626e+4	157.81	224.34	0.00
	Esférico	-0.201	0.82	1334e+4	156.86	228.97	40.99
	Gaussiano	0.007	1.02	1626e+4	157.81	224.34	0.00
1997	-	-	-	-	-	-	-
2002	Exponencial	-0.12	0.82	3296e+3	125.00	120.11	38.50
	Esférico	5.43e-14	0.91	5284e+3	125.50	104.38	0.00
	Gaussiano	-0.21	0.78	4180e+3	125.50	107.67	22.06
2006	-	-	-	-	-	-	-



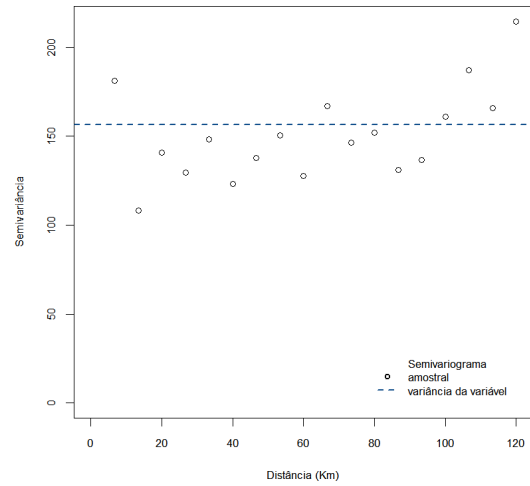
(a) Campanha de 1992



(b) Campanha de 1997



(c) Campanha de 2002



(d) Campanha de 2006

Figura 22: Variogramas amostrais das concentrações de Zn nos musgos observadas em cada campanha e respetivos modelos teóricos, estimados pelo método de mínimos quadrados.

Vanádio

Tabela 27: Estimativas dos parâmetros envolvidos nas estruturas do valor médio e da anisotropia das concentrações de V nos musgos observadas em cada campanha.

Campanha	Tendência						Anisotropia	
	$\hat{\beta}_0$	D.p.	$\hat{\beta}_1$	D.p.	$\hat{\beta}_2$	D.p.	$\hat{\psi}_A$	$\hat{\psi}_R$
2002	3.55	0.27	-	-	-	-	9.31e-09	1.00
2006	2.66	0.82	-	-	-	-	3.42	1.29

Tabela 28: Valores dos variogramas empíricos das concentrações de V nos musgos observadas em cada campanha, onde $|N(u)|$ representa o cardinal das diferentes classes de distâncias.

Campanha							
2002				2006			
u	$ N(u) $	$\hat{\gamma}(u)$	D.p.	u	$ N(u) $	$\hat{\gamma}(u)$	D.p.
0.00	29.00	11.16	25.40	0.00	41.00	29.89	58.88
20.83	338.00	10.58	26.95	20.00	438.00	21.65	46.89
41.67	521.00	11.63	27.63	40.00	613.00	15.57	41.71
62.50	615.00	12.09	31.11	60.00	748.00	16.12	38.59
83.33	721.00	11.25	29.41	80.00	727.00	16.78	42.83
104.17	786.00	11.25	29.38	100.00	609.00	16.58	39.76
125.00	859.00	10.07	24.04	120.00	467.00	15.48	42.30
145.83	832.00	10.20	28.98	140.00	322.00	16.98	38.80
166.67	799.00	12.76	28.63	160.00	158.00	16.10	36.04
187.50	718.00	12.16	27.71	180.00	41.00	7.54	23.24
208.33	652.00	10.84	24.91	200.00	18.00	8.32	58.88
229.17	566.00	10.97	24.05	220.00	3.00	8.16	46.89
250.00	255.00	13.19	27.39				

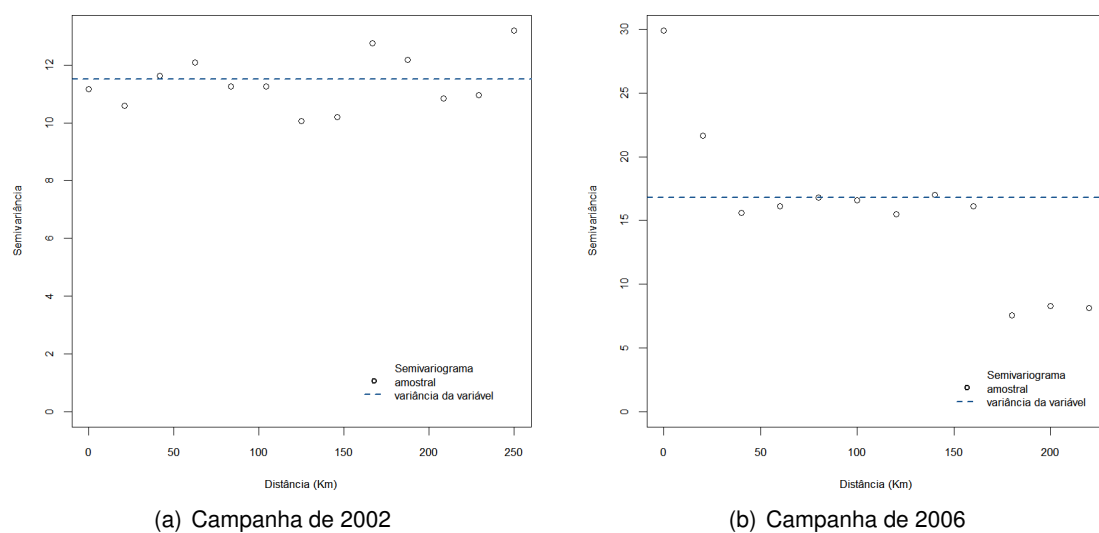


Figura 23: Variogramas amostrais das concentrações de V nos musgos observadas em cada campanha.

Mapas de Portugal continental com os valores preditos e com os desvios padrão associados aos elementos químicos

Arsénio

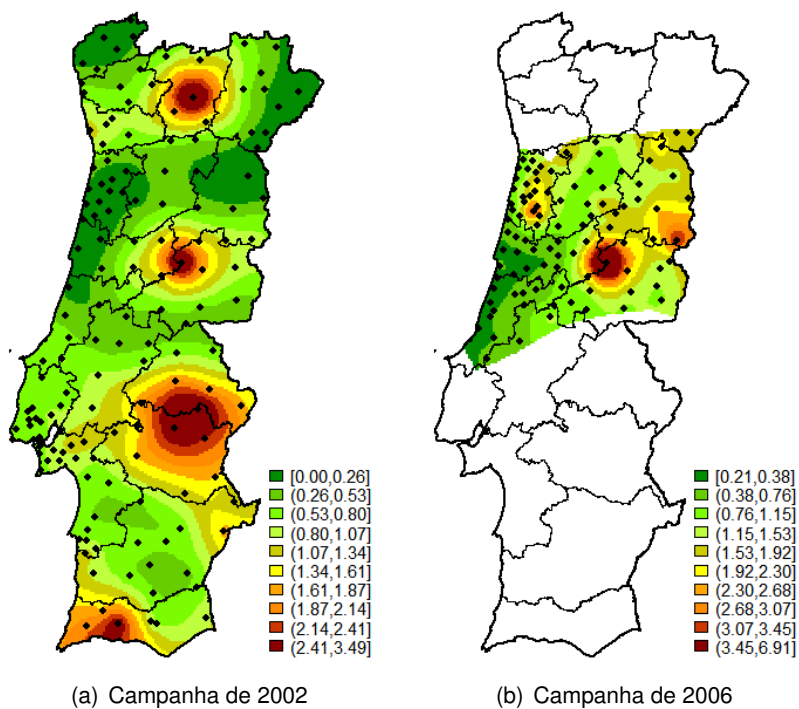


Figura 24: Superfícies preditas das concentrações de As nos musgos observadas em cada campanha, sem partilha de escala.

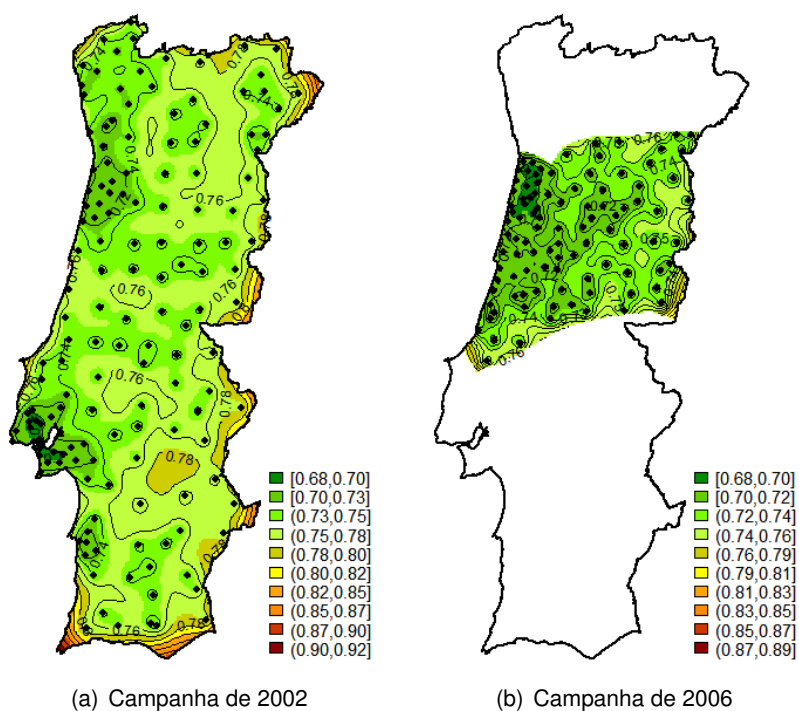


Figura 25: Desvios padrão associados a cada valor predito de As, cujas concentrações foram observadas nos musgos em cada campanha.

Cádmio

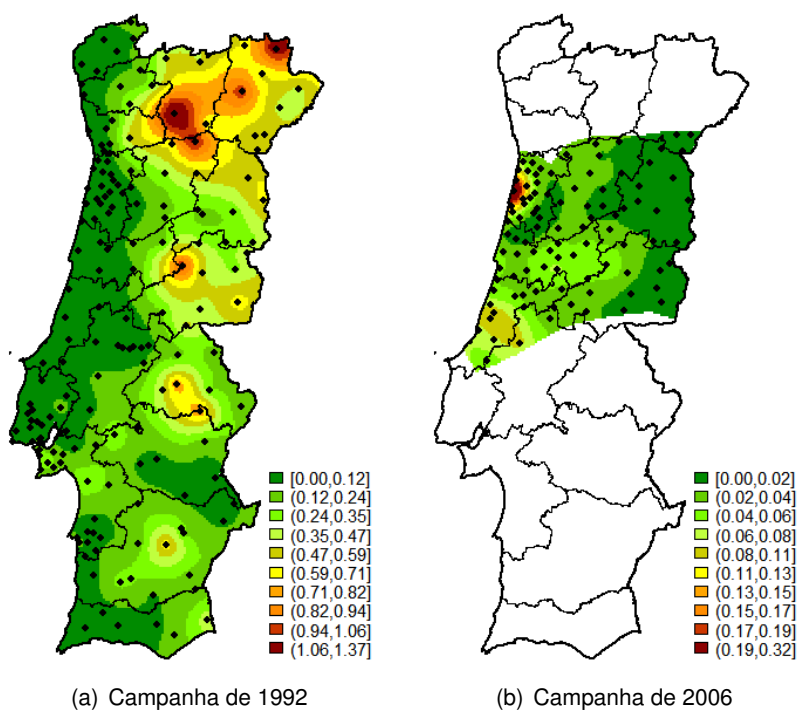


Figura 26: Superfícies preditas das concentrações de Cd nos musgos observadas em cada campanha, sem partilha de escala.

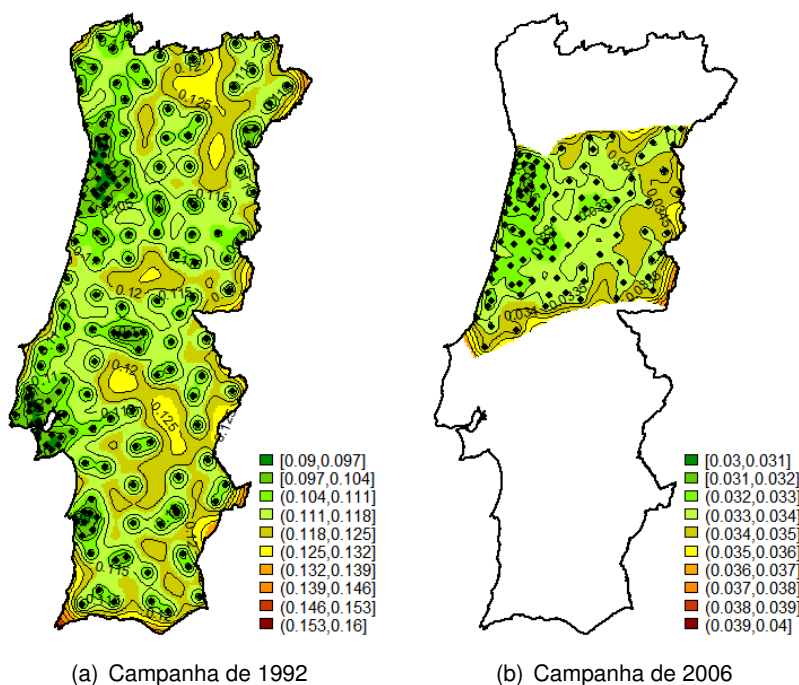


Figura 27: Desvios padrão associados a cada valor predito de Cd, cujas concentrações foram observadas nos musgos em cada campanha.

Crómio

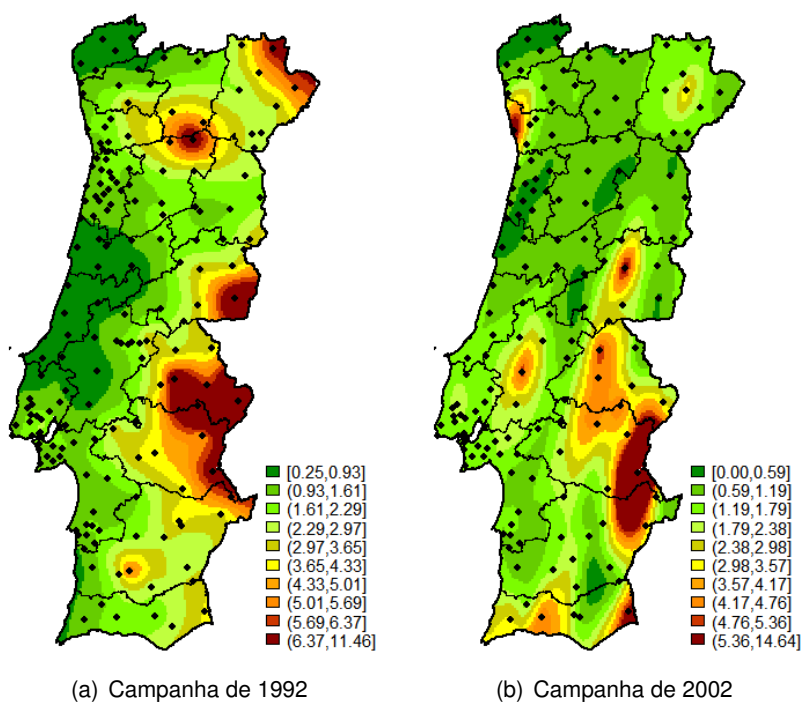


Figura 28: Superfícies preditas das concentrações de Cr nos musgos observadas em cada campanha, sem partilha de escala.

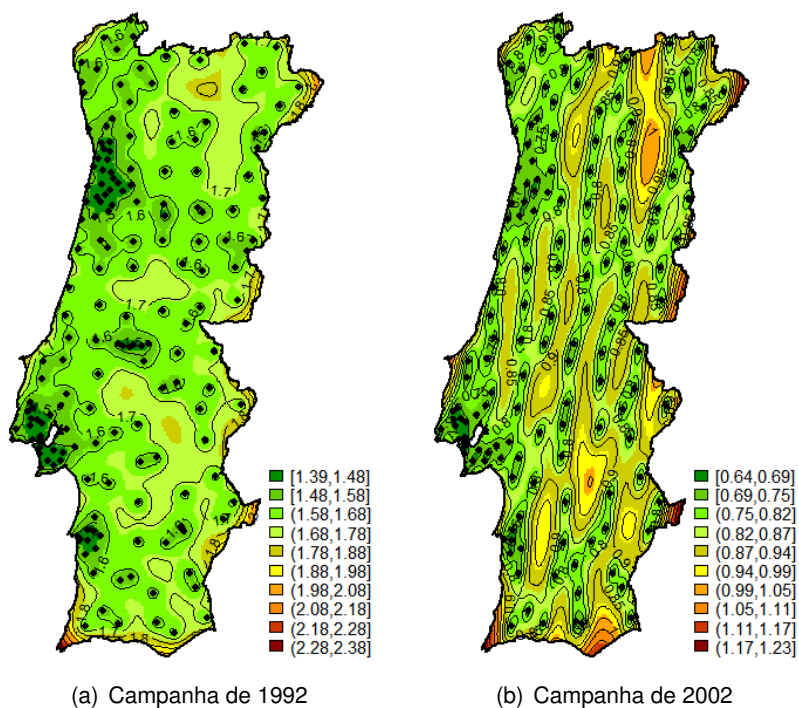


Figura 29: Desvios padrão associados a cada valor predito de Cr, cujas concentrações foram observadas nos musgos em cada campanha.

Cobre

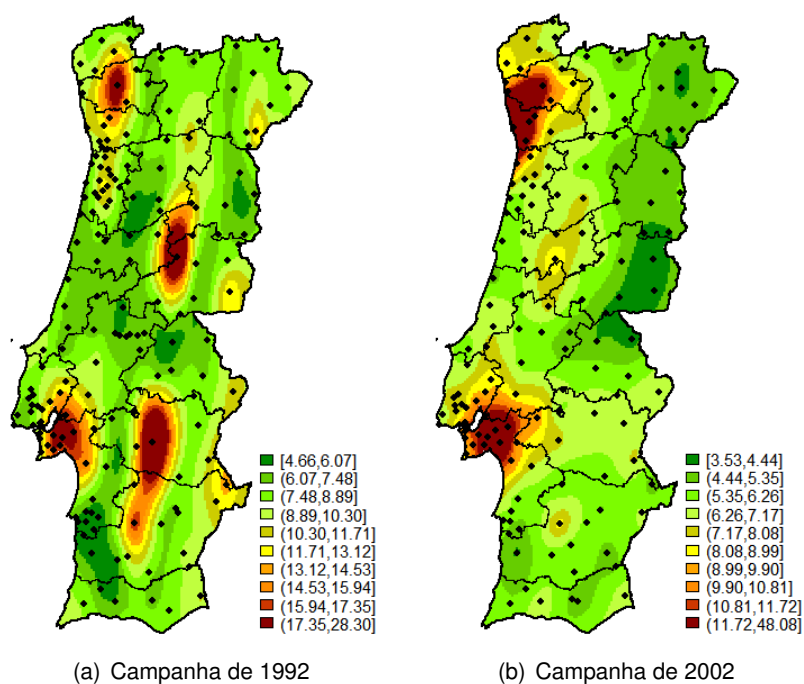


Figura 30: Superfícies previstas das concentrações de Cu nos musgos observadas em cada campanha, sem partilha de escala.

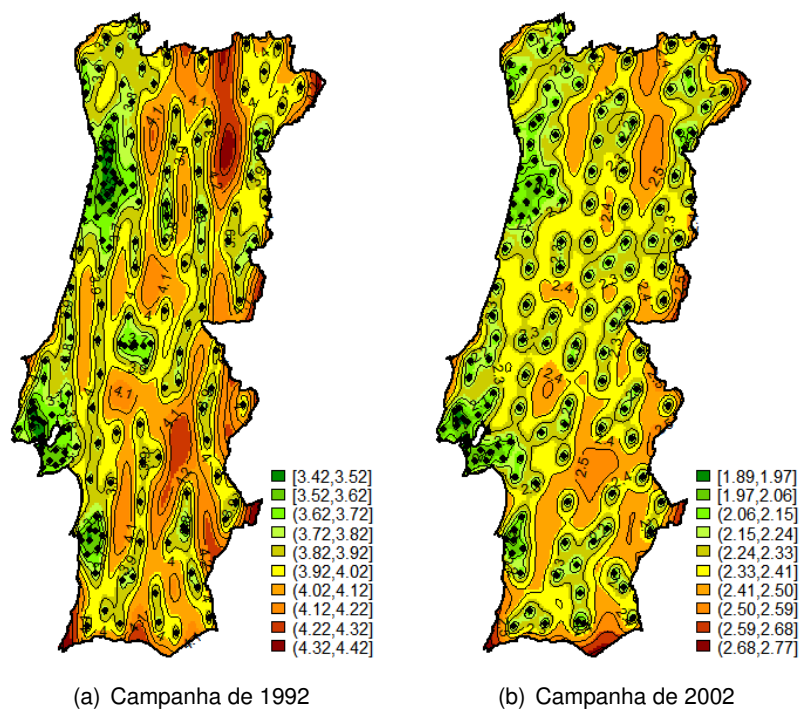


Figura 31: Desvios padrão associados a cada valor previsto de Cu, cujas concentrações foram observadas nos musgos em cada campanha.

Ferro

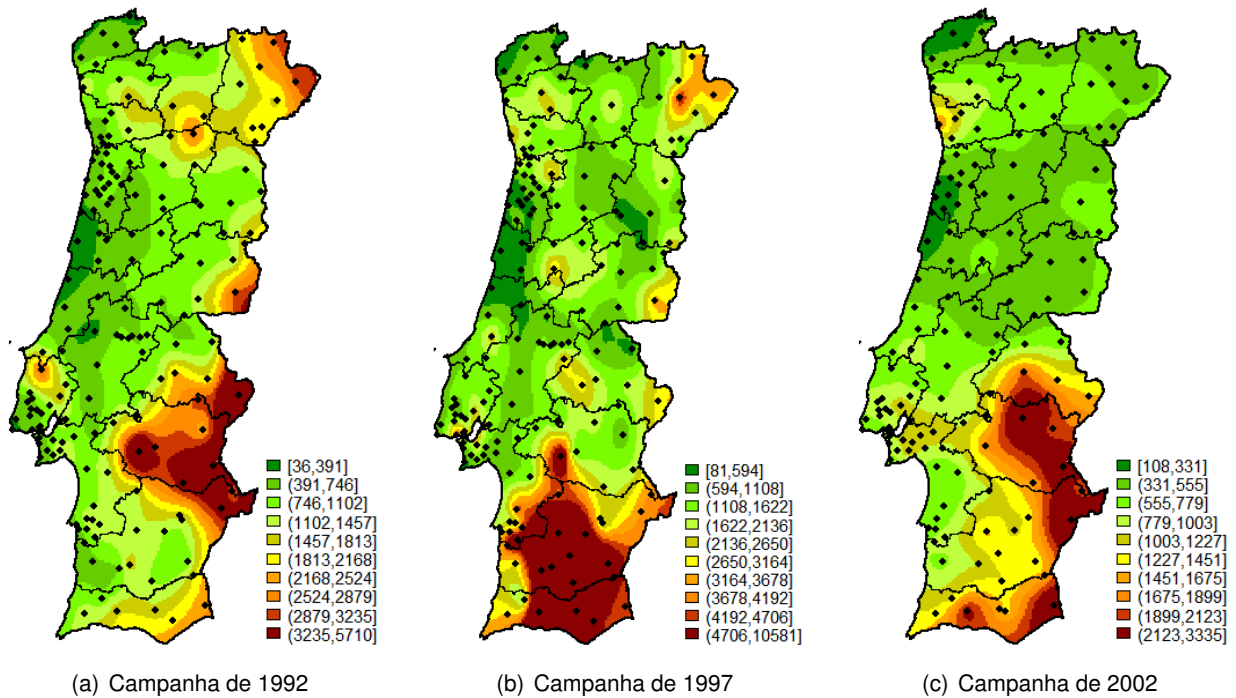


Figura 32: Superfícies previstas das concentrações de Fe nos musgos observadas em cada campanha, sem partilha de escala.

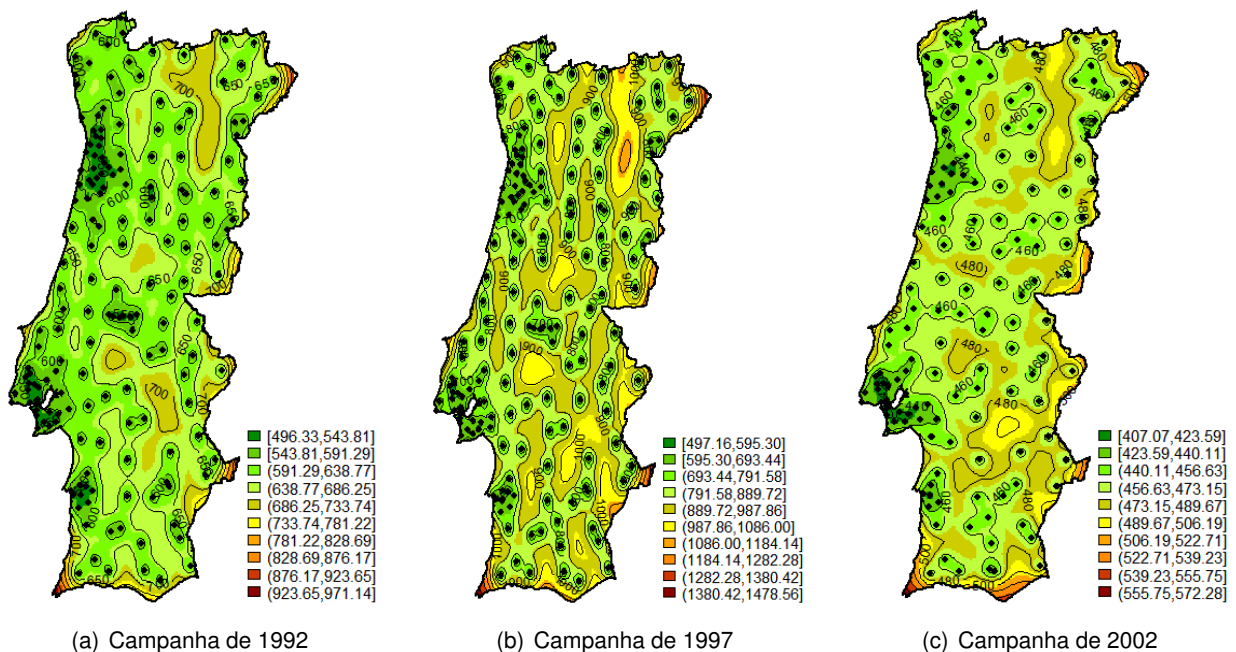


Figura 33: Desvios padrão associados a cada valor previsto de Fe, cujas concentrações foram observadas nos musgos em cada campanha.

Manganês

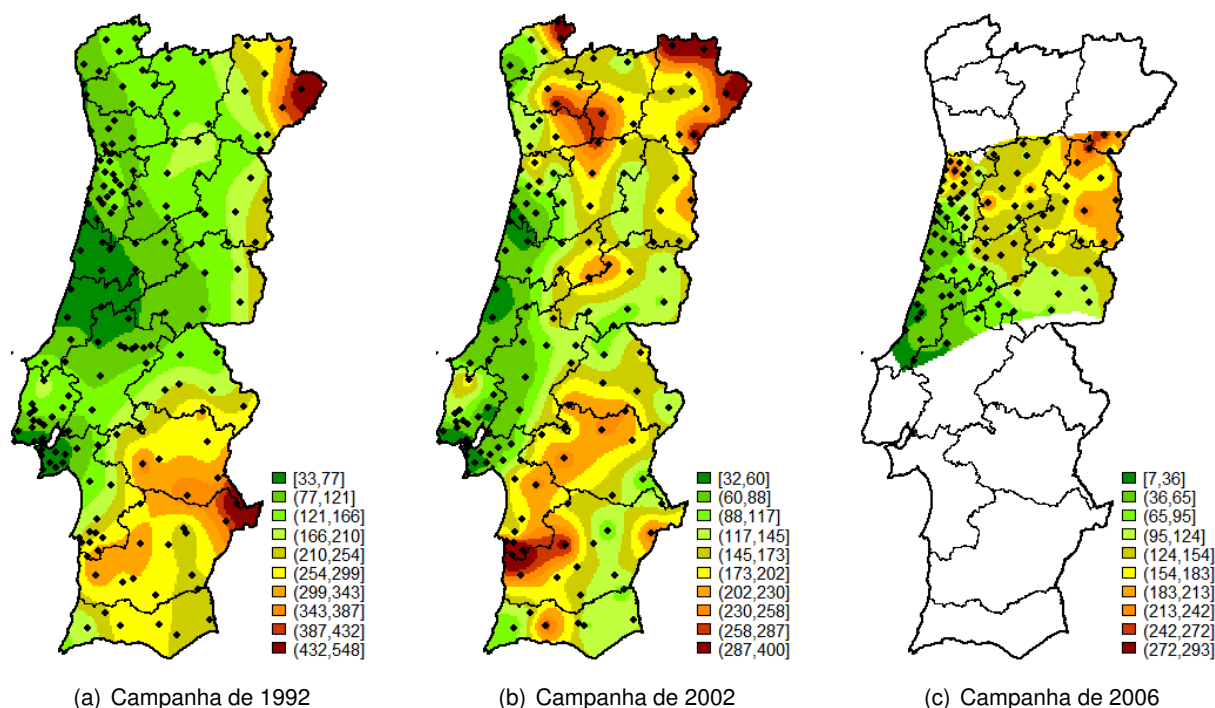


Figura 34: Superfícies preditas das concentrações de Mn nos musgos observadas em cada campanha, sem partilha de escala.

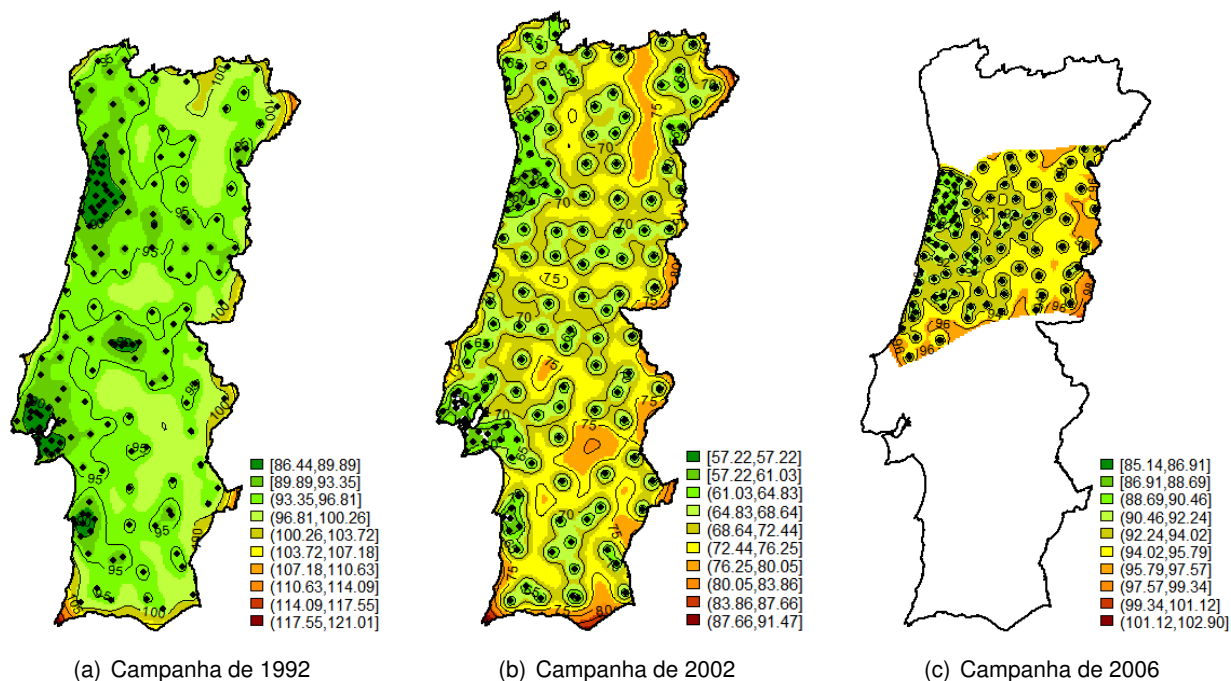


Figura 35: Desvios padrão associados a cada valor predito de Mn, cujas concentrações foram observadas nos musgos em cada campanha.

Níquel

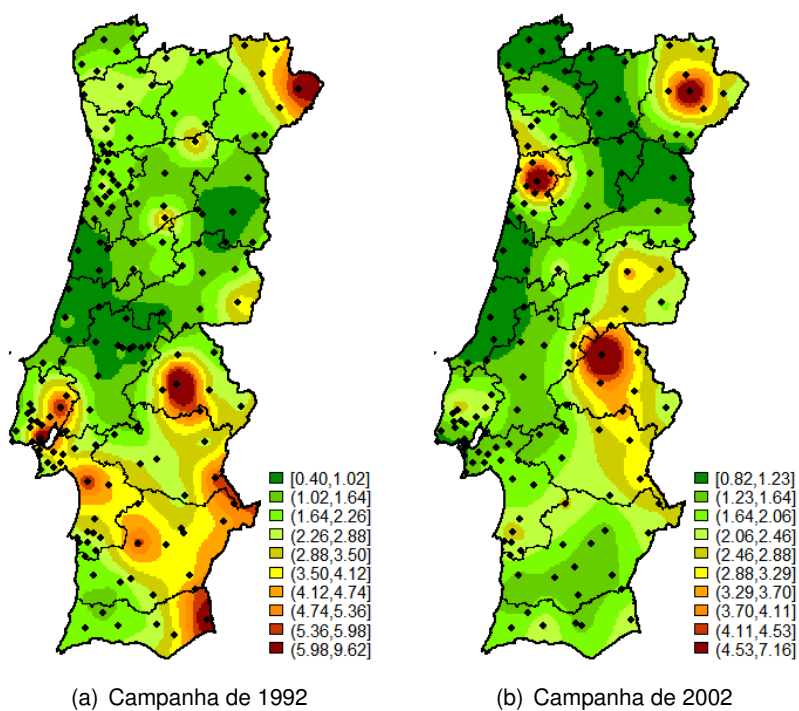


Figura 36: Superfícies preditas das concentrações de Ni nos musgos observadas em cada campanha, sem partilha de escala.

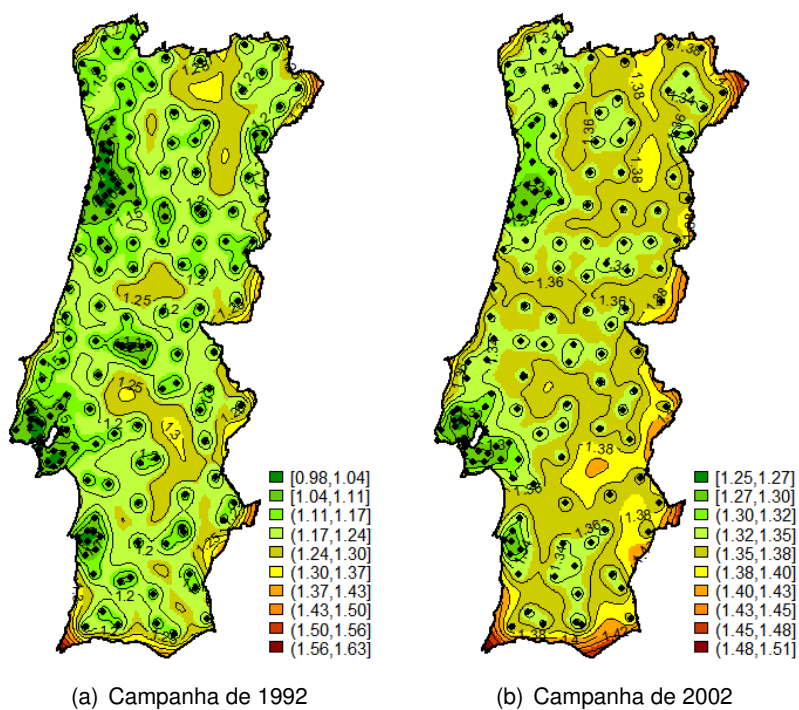


Figura 37: Desvios padrão associados a cada valor predito de Ni, cujas concentrações foram observadas nos musgos em cada campanha.

Chumbo

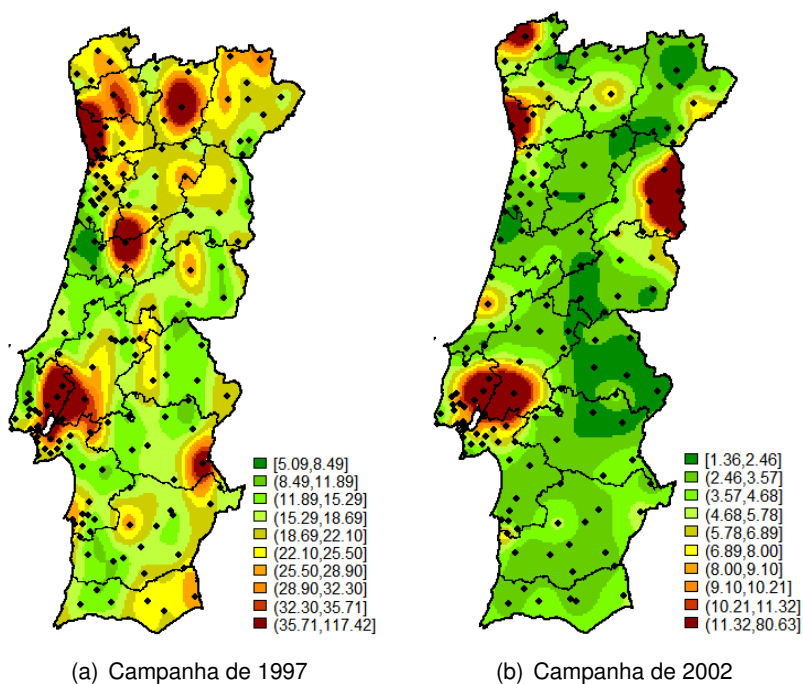


Figura 38: Superfícies previstas das concentrações de Pb nos musgos observadas em cada campanha, sem partilha de escala.

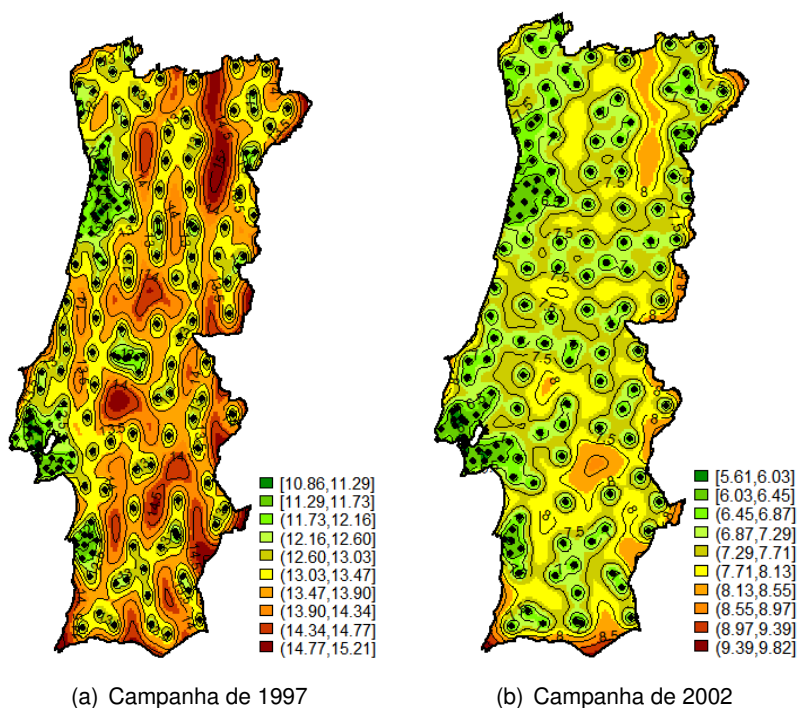


Figura 39: Desvios padrão associados a cada valor previsto de Pb, cujas concentrações foram observadas nos musgos em cada campanha.

Zinco

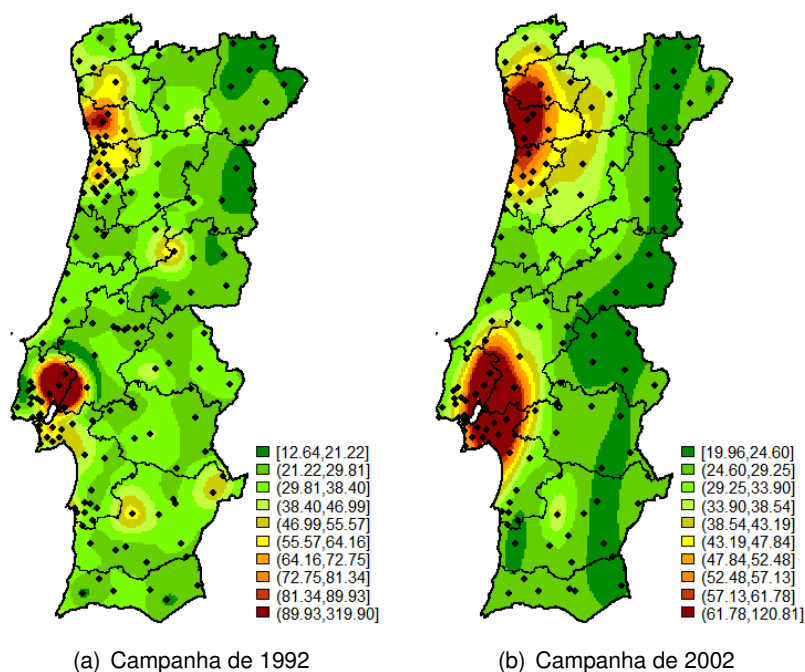


Figura 40: Superfícies preditas das concentrações de Zn nos musgos observadas em cada campanha, sem partilha de escala.

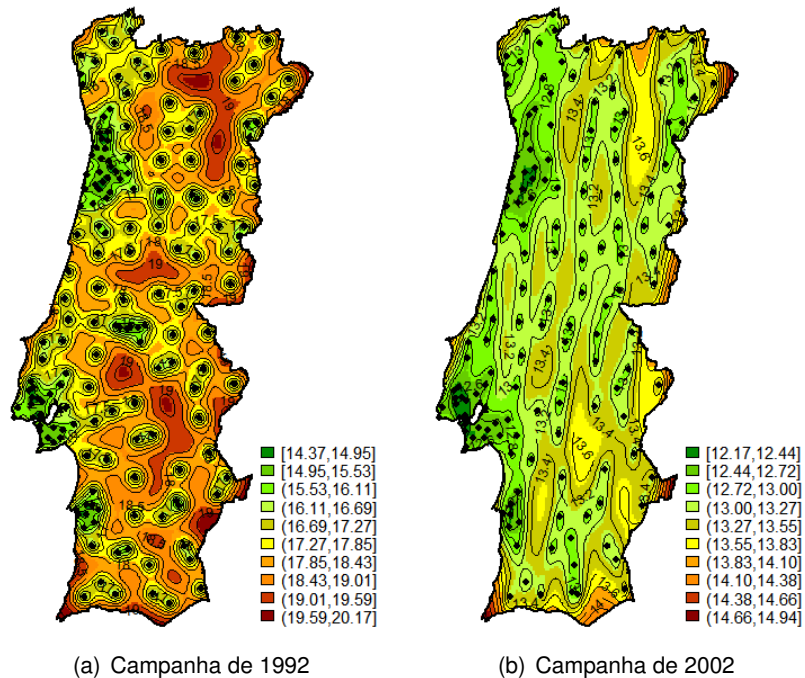


Figura 41: Desvios padrão associados a cada valor predito de Zn, cujas concentrações foram observadas nos musgos em cada campanha.

Anexo V

Diagramas de extremos e quartis das fontes de poluição

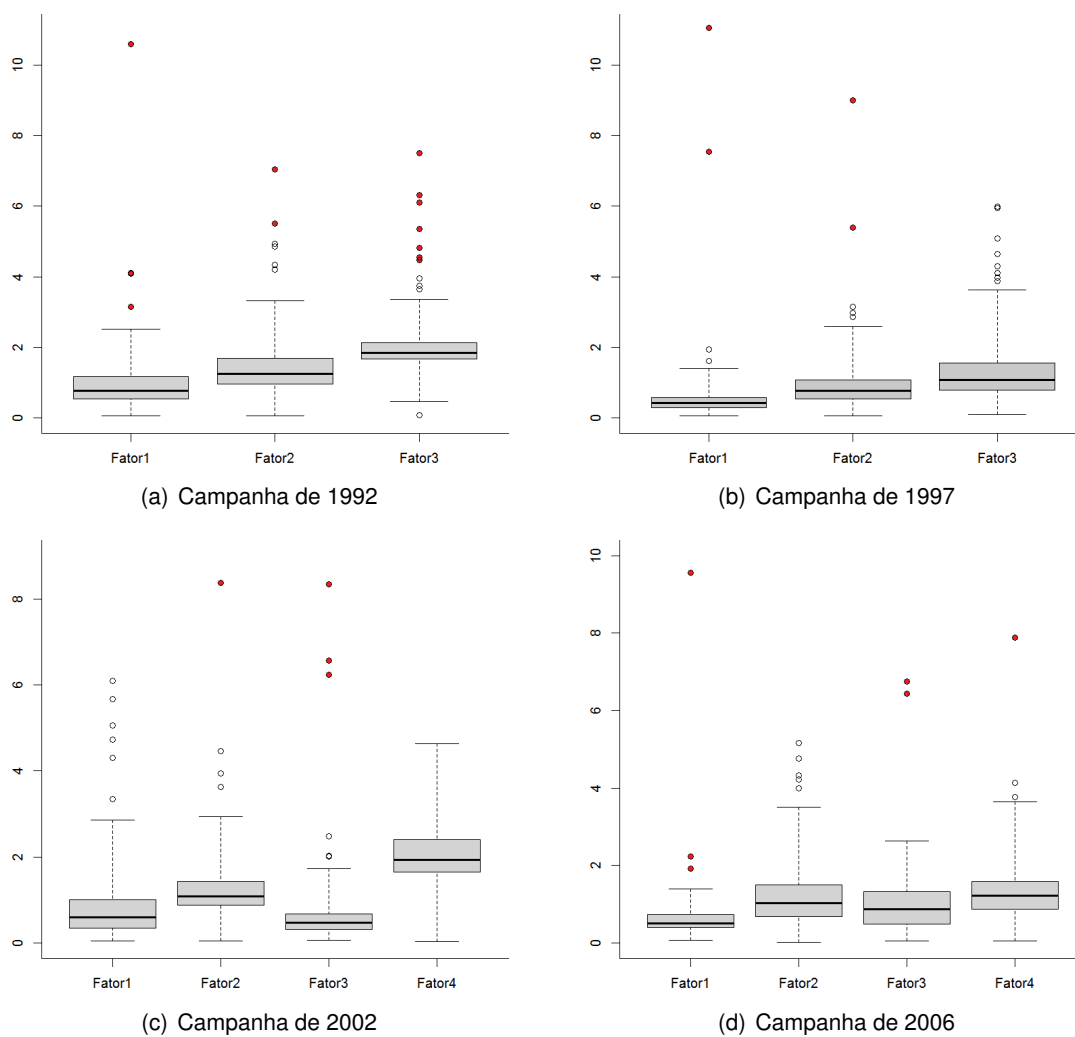


Figura 42: Diagramas de extremos e quartis das várias fontes de poluição subjacentes a cada campanha.

Anexo VI

Análise geoestatística das fontes de poluição

Campanha de 1992

Tabela 29: Estimativas dos parâmetros envolvidos nas estruturas do valor médio e da anisotropia das fontes de poluição subjacentes à campanha de 1992.

Fator	Tendência						Anisotropia	
	$\hat{\beta}_0$	D.p.	$\hat{\beta}_1$	D.p.	$\hat{\beta}_2$	D.p.	$\hat{\psi}_A$	$\hat{\psi}_R$
Ação antropogénica	-0.13	0.04	-0.002	0.001	0.001	0.000	0.33	1.62
Componente geológica	1.52	0.05	0.006	0.001	-0.001	0.000	2.57e-09	1.00
Presença de minas e agentes erosivos	1.97	0.04	0.002	0.002	-0.000	0.000	0.00	2.77

Tabela 30: Valores dos variogramas empíricos das fontes de poluição subjacentes à campanha de 1992, onde $|N(u)|$ representa o cardinal das diferentes classes de distâncias.

Fator							
Ação antropogénica				Componente geológica			
u	$ N(u) $	$\hat{\gamma}(u)$	D.p.	u	$ N(u) $	$\hat{\gamma}(u)$	D.p.
0.00	114.00	0.19	0.29	0.00	63.00	0.18	0.37
20.83	701.00	0.23	0.35	20.00	494.00	0.25	0.83
41.67	932.00	0.25	0.36	40.00	653.00	0.29	0.98
62.50	1201.00	0.24	0.35	60.00	759.00	0.38	1.23
83.33	1365.00	0.24	0.35	80.00	869.00	0.39	1.18
104.17	1450.00	0.24	0.39	100.00	983.00	0.35	0.94
125.00	1541.00	0.25	0.38	120.00	1061.00	0.40	1.01
145.83	1506.00	0.25	0.36	140.00	1082.00	0.44	1.14
166.67	1272.00	0.24	0.37	160.00	1013.00	0.49	1.30
187.50	1030.00	0.24	0.39	180.00	895.00	0.51	1.31
208.33	823.00	0.23	0.39	200.00	777.00	0.50	1.32
229.17	695.00	0.24	0.38	220.00	732.00	0.45	1.27
250.00	267.00	0.27	0.44	240.00	724.00	0.38	1.02
				260.00	641.00	0.43	1.08
				280.00	523.00	0.53	1.16
				300.00	524.00	0.43	0.93
				320.00	260.00	0.51	1.29

Fator			
Presença de minas e agentes erosivos			
u	$ N(u) $	$\hat{\gamma}(u)$	D.p.
0.00	81.00	0.11	0.27
12.86	546.00	0.20	0.54
25.71	763.00	0.22	0.45
38.57	852.00	0.23	0.50
51.43	1026.00	0.23	0.45
64.29	1095.00	0.26	0.50
77.14	1156.00	0.30	0.52
90.00	1246.00	0.30	0.54
102.86	1250.00	0.25	0.48
115.71	1203.00	0.25	0.44
128.57	1115.00	0.27	0.47
141.43	945.00	0.26	0.44
154.29	792.00	0.26	0.41
167.14	499.00	0.27	0.45
180.00	194.00	0.23	0.42

Tabela 31: Estimativas dos parâmetros envolvidos nas estruturas da associação espacial dos modelos adotados para as fontes de poluição subjacentes à campanha de 1992.

Fator	Modelo	Parâmetros					
		EM	MEEQ	SQR	$\hat{\tau}^2$	$\hat{\sigma}^2$	$\hat{\phi}$
Ação antropogénica	Exponencial	-0.001	0.93	0.58	0.20	0.04	12.89
	Esférico	-0.001	0.95	0.64	0.20	0.04	20.00
	Gaussiano	-0.002	0.92	0.56	0.20	0.04	17.68
Componente geológica	Exponencial	0.001	1.31	21.19	0.18	0.30	70.16
	Esférico	0.001	1.19	20.68	0.18	0.29	174.71
	Gaussiano	0.002	1.73	25.32	0.18	0.27	70.83
Presença de minas e agentes erosivos	Exponencial	-0.002	1.25	4.66	0.11	0.16	21.87
	Esférico	-0.001	1.60	5.66	0.13	0.14	75.92
	Gaussiano	-0.002	1.86	7.08	0.08	0.18	14.71

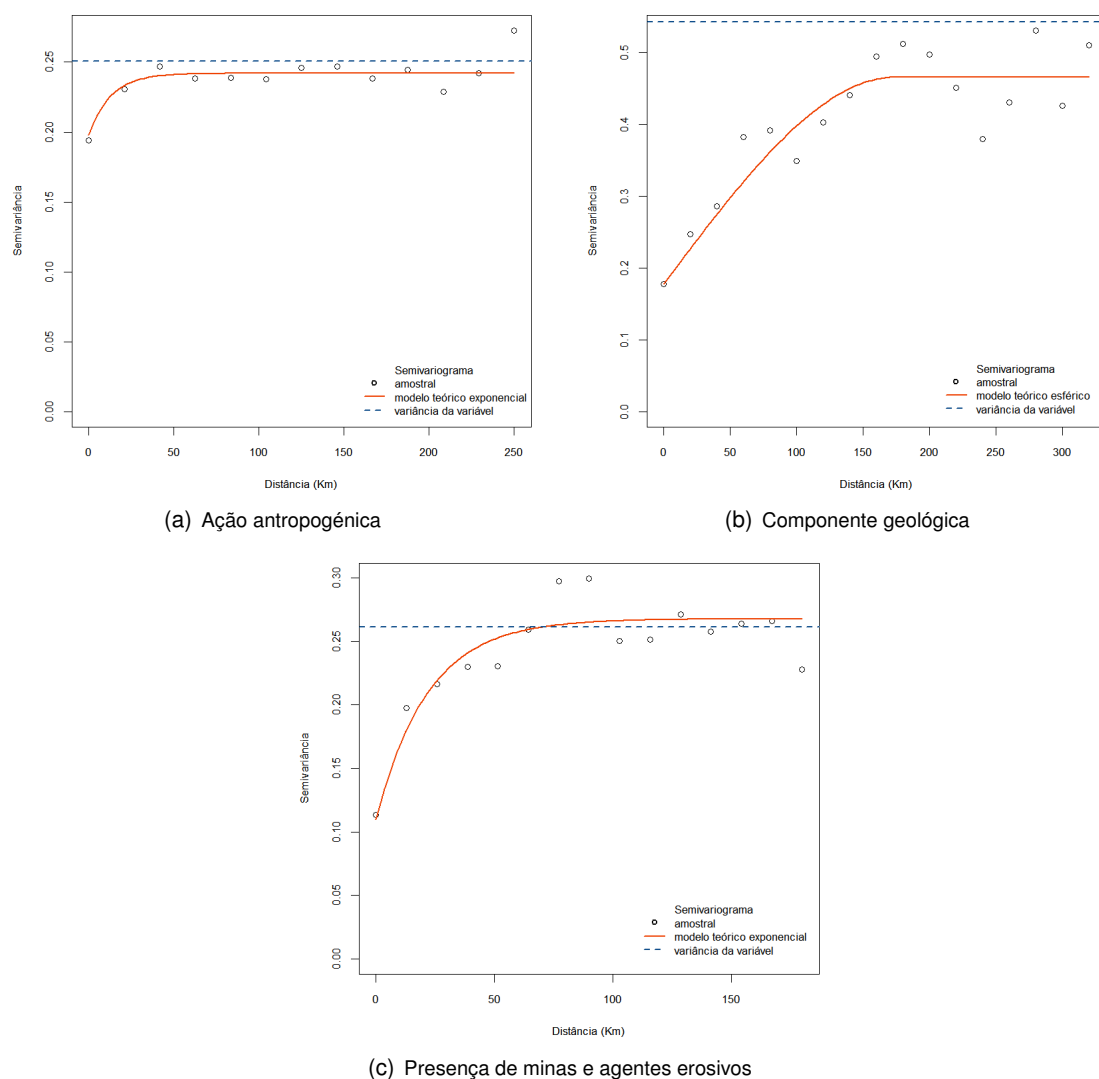


Figura 43: Variogramas amostrais das fontes de poluição subjacentes à campanha de 1992 e respetivos modelos teóricos, estimados pelo método de mínimos quadrados.

Campanha de 1997

Tabela 32: Estimativas dos parâmetros envolvidos nas estruturas do valor médio e da anisotropia das fontes de poluição subjacentes à campanha de 1997.

Fator	Tendência						Anisotropia	
	$\hat{\beta}_0$	D.p.	$\hat{\beta}_1$	D.p.	$\hat{\beta}_2$	D.p.	$\hat{\psi}_A$	$\hat{\psi}_R$
Emissões da queima de combustíveis fósseis	0.49	0.02	-	-	-	-	0.00	2.18
Ação antropogénica	0.85	0.04	-	-	-	-	0.89	3.73
Componente geológica	1.46	0.06	0.004	0.001	-0.004	0.000	0.00	2.77

Tabela 33: Valores dos variogramas empíricos das fontes de poluição subjacentes à campanha de 1997, onde $|N(u)|$ representa o cardinal das diferentes classes de distâncias.

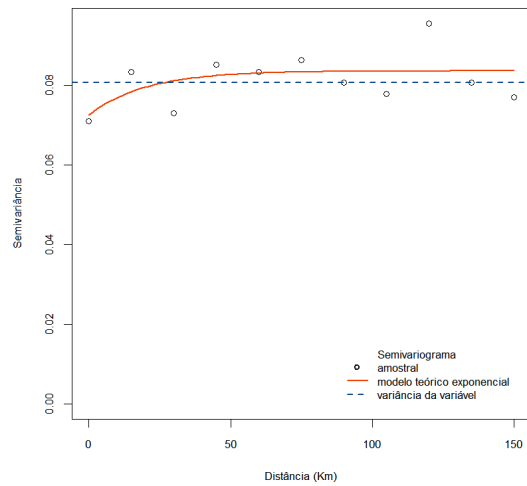
Fator							
Emissões da queima de combustíveis fósseis				Ação antropogénica			
u	$ N(u) $	$\hat{\gamma}(u)$	D.p.	u	$ N(u) $	$\hat{\gamma}(u)$	D.p.
0.00	90.00	0.07	0.12	0.00	228.00	0.24	0.44
15.00	594.00	0.08	0.13	20.00	1085.00	0.25	0.43
30.00	860.00	0.07	0.14	40.00	1548.00	0.29	0.52
45.00	956.00	0.09	0.16	60.00	1548.00	0.27	0.50
60.00	1183.00	0.08	0.14	80.00	1288.00	0.28	0.53
75.00	1261.00	0.09	0.15	100.00	1209.00	0.27	0.49
90.00	1379.00	0.08	0.14	120.00	1240.00	0.26	0.49
105.00	1379.00	0.08	0.14	140.00	1134.00	0.31	0.53
120.00	1448.00	0.10	0.16	160.00	1046.00	0.33	0.58
135.00	1401.00	0.08	0.14	180.00	853.00	0.24	0.44
150.00	600.00	0.08	0.13	200.00	394.00	0.27	0.53

Fator			
Componente geológica			
u	$ N(u) $	$\hat{\gamma}(u)$	D.p.
0.00	85.00	0.32	0.59
10.71	579.00	0.38	0.93
21.43	809.00	0.38	0.77
32.14	1015.00	0.51	1.17
42.86	1050.00	0.67	1.48
53.57	1281.00	0.67	1.41
64.29	1389.00	0.75	1.50
75.00	1525.00	0.75	1.44
85.71	1462.00	0.83	1.43
96.43	1237.00	1.01	1.65
107.14	1059.00	0.93	1.54
117.86	1027.00	0.86	1.42
128.57	809.00	0.79	1.39
139.29	680.00	0.59	1.03
150.00	288.00	0.55	1.11

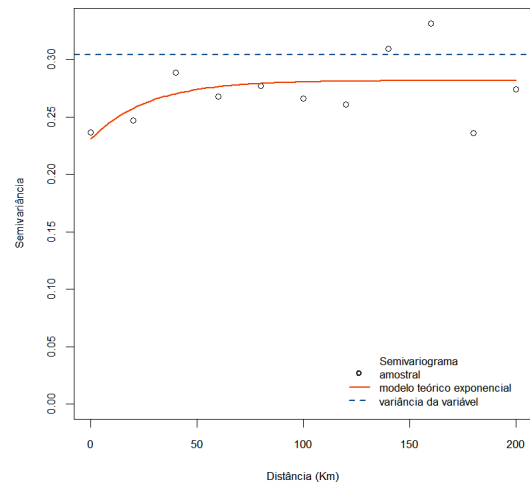
Tabela 34: Estimativas dos parâmetros envolvidos nas estruturas da associação espacial dos modelos adotados para as fontes de poluição subjacentes à campanha de 1997.

Fator	Modelo	Parâmetros					
		EM	MEEQ	SQR	$\hat{\tau}^2$	$\hat{\sigma}^2$	$\hat{\phi}$
Emissões da queima de combustíveis fósseis	Exponencial	6.13e-05	0.96	0.39	0.07	0.01	20.02
	Esférico	-8.67e-05	0.96	0.41	0.07	0.01	20.02
	Gaussiano	0.001	0.96	0.38	0.08	0.01	40.82

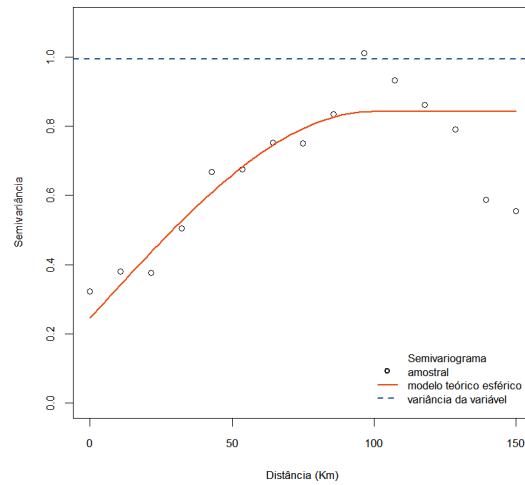
Fator	Modelo	Parâmetros					
		EM	MEEQ	SQR	$\hat{\tau}^2$	$\hat{\sigma}^2$	$\hat{\phi}$
Ação antropogénica	Exponencial	-0.001	0.91	6.82	0.23	0.05	26.52
	Esférico	-0.001	0.83	7.52	0.24	0.04	20.00
	Gaussiano	-0.001	0.85	6.63	0.23	0.05	23.26
Componente geológica	Exponencial	-0.005	1.82	153.27	0.11	0.75	34.53
	Esférico	-0.002	1.26	125.71	0.25	0.60	98.94
	Gaussiano	-0.001	1.28	127.82	0.32	0.52	46.24



(a) Emissões da queima de combustíveis fósseis



(b) Ação antropogénica



(c) Componente geológica

Figura 44: Variogramas amostrais das fontes de poluição subjacentes à campanha de 1997 e respetivos modelos teóricos, estimados pelo método de mínimos quadrados.

Campanha de 2002

Tabela 35: Estimativas dos parâmetros envolvidos nas estruturas do valor médio e da anisotropia das fontes de poluição subjacentes à campanha de 2002.

Fator	Tendência						Anisotropia	
	$\hat{\beta}_0$	D.p.	$\hat{\beta}_1$	D.p.	$\hat{\beta}_2$	D.p.	$\hat{\psi}_A$	$\hat{\psi}_R$
Contaminação por partículas do solo	0.93	0.07	0.004	0.001	-0.003	0.001	0.06	5.82
Ação antropogénica	1.22	0.05	-	-	-	-	0.01	1.80
Contaminação pela emissão de Ni	0.58	0.03	-	-	-	-	5.51e-12	1.00
Contaminação pela emissão de Mn	2.11	0.05	-	-	-	-	0.00	1.10

Tabela 36: Valores dos variogramas empíricos das fontes de poluição subjacentes à campanha de 2002, onde $|N(u)|$ representa o cardinal das diferentes classes de distâncias.

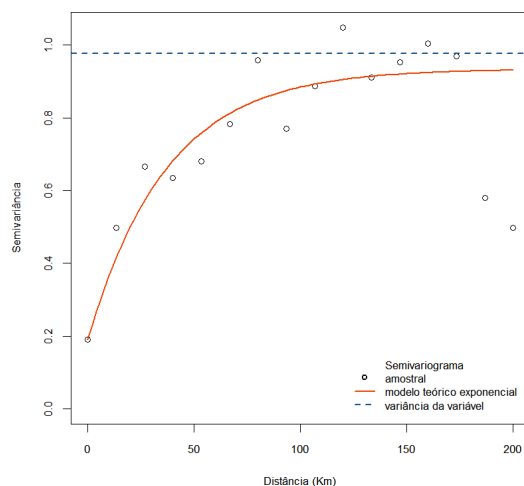
Fator							
Contaminação por partículas do solo				Ação antropogénica			
u	$ N(u) $	$\hat{\gamma}(u)$	D.p.	u	$ N(u) $	$\hat{\gamma}(u)$	D.p.
0.00	110.00	0.19	0.49	0.00	68.00	0.27	0.66
13.33	648.00	0.50	1.46	20.83	559.00	0.35	0.85
26.67	1041.00	0.67	1.96	41.67	810.00	0.41	0.91
40.00	1153.00	0.63	1.73	62.50	1013.00	0.34	0.84
53.33	1520.00	0.68	1.84	83.33	1156.00	0.35	0.89
66.67	1319.00	0.78	1.89	104.17	1174.00	0.42	1.03
80.00	1199.00	0.96	2.21	125.00	1303.00	0.44	0.95
93.33	952.00	0.77	1.78	145.83	1204.00	0.39	0.88
106.67	795.00	0.89	2.07	166.67	1001.00	0.37	0.89
120.00	656.00	1.05	2.40	187.50	873.00	0.37	0.90
133.33	589.00	0.91	2.09	208.33	650.00	0.36	0.83
146.67	447.00	0.95	2.11	229.17	465.00	0.38	0.79
160.00	292.00	1.00	2.08	250.00	188.00	0.55	1.29
173.33	215.00	0.97	2.35				
186.67	134.00	0.58	1.51				
200.00	38.00	0.50	1.75				

Fator							
Contaminação pela emissão de Ni				Contaminação pela emissão de Mn			
u	$ N(u) $	$\hat{\gamma}(u)$	D.p.	u	$ N(u) $	$\hat{\gamma}(u)$	D.p.
0.00	28.00	0.12	0.28	0.00	10.00	0.11	0.13
21.54	342.00	0.13	0.27	15.62	201.00	0.21	0.42
43.08	531.00	0.16	0.35	31.25	370.00	0.27	0.46
64.62	617.00	0.15	0.32	46.88	461.00	0.36	0.62
86.15	752.00	0.16	0.37	62.50	486.00	0.37	0.54
107.69	771.00	0.17	0.35	78.12	569.00	0.48	0.65
129.23	853.00	0.17	0.37	93.75	619.00	0.47	0.73
150.77	817.00	0.17	0.34	109.38	633.00	0.41	0.59

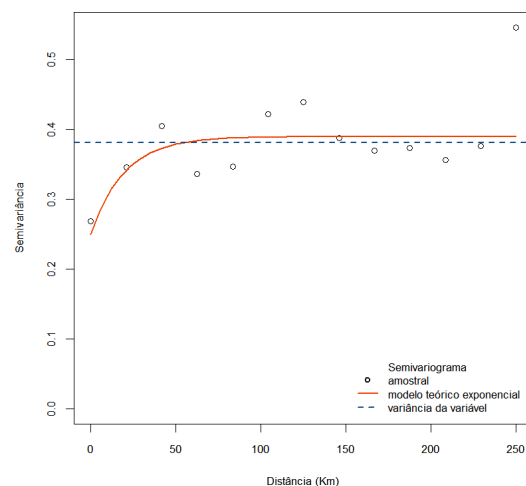
Fator							
Contaminação pela emissão de Ni				Contaminação pela emissão de Mn			
u	$ N(u) $	$\hat{\gamma}(u)$	D.p.	u	$ N(u) $	$\hat{\gamma}(u)$	D.p.
172.31	774.00	0.17	0.34	125.00	664.00	0.43	0.67
193.85	675.00	0.19	0.40	140.62	682.00	0.43	0.69
215.38	617.00	0.18	0.41	156.25	667.00	0.41	0.59
236.92	542.00	0.16	0.35	171.88	624.00	0.49	0.74
258.46	483.00	0.15	0.31	187.50	566.00	0.47	0.70
280.00	220.00	0.14	0.29	203.12	507.00	0.42	0.67
				218.75	472.00	0.38	0.62
				234.38	398.00	0.42	0.62
				250.00	192.00	0.45	0.64

Tabela 37: Estimativas dos parâmetros envolvidos nas estruturas da associação espacial dos modelos adotados para as fontes de poluição subjacentes à campanha de 2002.

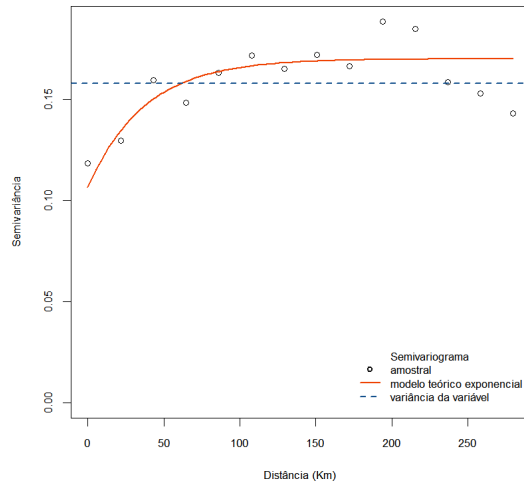
Fator	Modelo	Parâmetros					
		EM	MEEQ	SQR	$\hat{\tau}^2$	$\hat{\sigma}^2$	$\hat{\phi}$
Contaminação por partículas do solo	Exponencial	0.003	1.69	90.96	0.19	0.74	36.91
	Esférico	0.006	2.06	116.71	0.19	0.71	87.19
	Gaussiano	0.001	2.73	162.91	0.19	0.67	32.81
Ação antropogénica	Exponencial	-0.004	0.78	15.72	0.25	0.14	19.52
	Esférico	-0.006	0.75	15.71	0.25	0.14	40.65
	Gaussiano	-0.008	0.72	15.74	0.25	0.14	18.49
Contaminação pela emissão de Ni	Exponencial	-0.001	1.05	0.91	0.11	0.06	37.02
	Esférico	-0.001	1.05	0.91	0.12	0.05	122.33
	Gaussiano	-0.001	1.06	0.93	0.13	0.04	62.08
Contaminação pela emissão de Mn	Exponencial	0.008	1.53	10.03	0.00	0.44	26.31
	Esférico	0.004	1.22	8.26	0.11	0.33	85.75
	Gaussiano	0.004	1.58	8.67	0.16	0.27	42.18



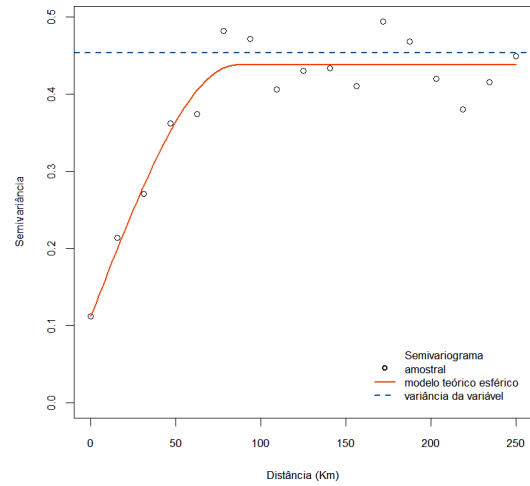
(a) Contaminação por partículas do solo



(b) Ação antropogénica



(c) Contaminação pela emissão de Ni



(d) Contaminação pela emissão de Mn

Figura 45: Variogramas amostrais das fontes de poluição subjacentes à campanha de 2002 e respetivos modelos teóricos, estimados pelo método de mínimos quadrados.

Campanha de 2006

Tabela 38: Estimativas dos parâmetros envolvidos nas estruturas do valor médio e da anisotropia das fontes de poluição subjacentes à campanha de 2006.

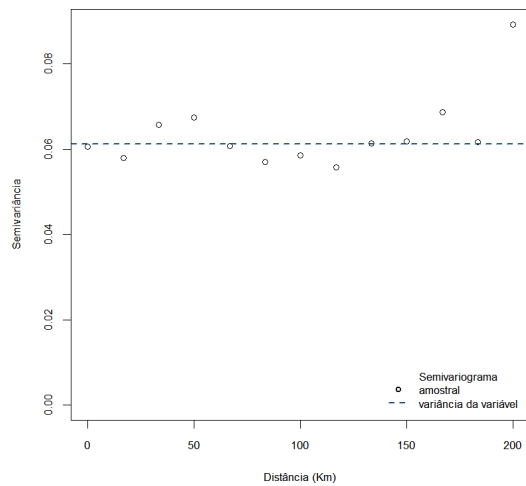
Fator	Tendência						Anisotropia	
	$\hat{\beta}_0$	D.p.	$\hat{\beta}_1$	D.p.	$\hat{\beta}_2$	D.p.	$\hat{\psi}_A$	$\hat{\psi}_R$
Presença de minas e agentes erosivos	0.60	0.05	-	-	-	-	2.21e-10	1.00
Contaminação por partículas do solo	1.02	0.19	-	-	-	-	0.084	6.61
Ação antropogénica	0.58	0.10	-0.003	0.001	0.005	0.001	5.71e-08	1.00
Contaminação pela emissão de Hg	1.30	0.15	-	-	-	-	0.73	7.11

Tabela 39: Valores dos variogramas empíricos das fontes de poluição subjacentes à campanha de 2006, onde $|N(u)|$ representa o cardinal das diferentes classes de distâncias.

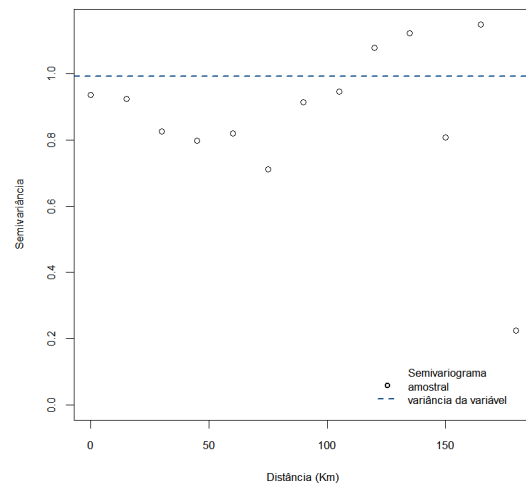
Fator							
Presença de minas e agentes erosivos				Contaminação por partículas do solo			
u	$ N(u) $	$\hat{\gamma}(u)$	D.p.	u	$ N(u) $	$\hat{\gamma}(u)$	D.p.
0.00	14.00	0.06	0.08	0.00	196.00	0.94	1.95
16.67	283.00	0.06	0.10	15.00	944.00	0.92	1.89
33.33	433.00	0.07	0.10	30.00	807.00	0.83	1.71
50.00	535.00	0.07	0.11	45.00	562.00	0.80	1.63
66.67	615.00	0.06	0.10	60.00	519.00	0.82	1.64
83.33	604.00	0.06	0.09	75.00	434.00	0.71	1.55
100.00	540.00	0.06	0.10	90.00	360.00	0.91	1.75
116.67	467.00	0.06	0.10	105.00	295.00	0.95	1.70
133.33	387.00	0.06	0.11	120.00	255.00	1.08	2.02
150.00	296.00	0.06	0.11	135.00	215.00	1.12	2.05
166.67	141.00	0.07	0.13	150.00	126.00	0.81	1.73

Fator							
Presença de minas e agentes erosivos				Contaminação por partículas do solo			
u	$ N(u) $	$\hat{\gamma}(u)$	D.p.	u	$ N(u) $	$\hat{\gamma}(u)$	D.p.
183.33	72.00	0.06	0.12	165.00	38.00	1.15	1.96
200.00	23.00	0.09	0.15	180.00	2.00	0.22	0.22

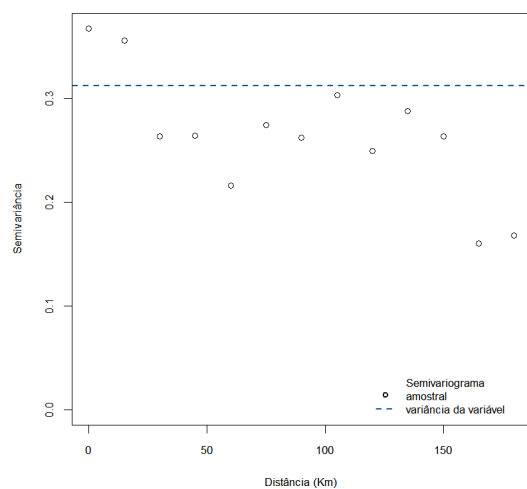
Fator							
Ação antropogénica				Contaminação pela emissão de Hg			
u	$ N(u) $	$\hat{\gamma}(u)$	D.p.	u	$ N(u) $	$\hat{\gamma}(u)$	D.p.
0.00	8.00	0.37	0.42	0.00	186.00	0.54	1.02
15.00	226.00	0.36	0.56	15.00	907.00	0.56	1.08
30.00	389.00	0.26	0.42	30.00	925.00	0.54	1.01
45.00	461.00	0.26	0.38	45.00	755.00	0.56	1.05
60.00	535.00	0.22	0.29	60.00	618.00	0.58	1.04
75.00	544.00	0.27	0.38	75.00	469.00	0.58	1.09
90.00	537.00	0.26	0.38	90.00	328.00	0.50	0.94
105.00	476.00	0.30	0.40	105.00	204.00	0.54	0.93
120.00	406.00	0.25	0.33	120.00	116.00	0.74	1.18
135.00	365.00	0.29	0.40	135.00	88.00	0.51	1.04
150.00	282.00	0.26	0.37	150.00	45.00	0.47	0.76
165.00	142.00	0.16	0.22	165.00	13.00	0.52	0.66
180.00	48.00	0.17	0.34	180.00	2.00	0.56	0.60



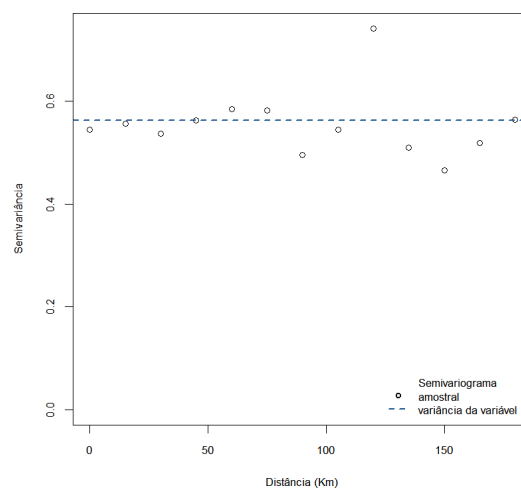
(a) Presença de minas e agentes erosivos



(b) Contaminação por partículas do solo



(c) Ação antropogénica



(d) Contaminação pela emissão de Hg

Figura 46: Variogramas amostrais das fontes de poluição subjacentes à campanha de 2006.

Mapas de Portugal continental com os desvios padrão associados às fontes de poluição

Campanha de 1992

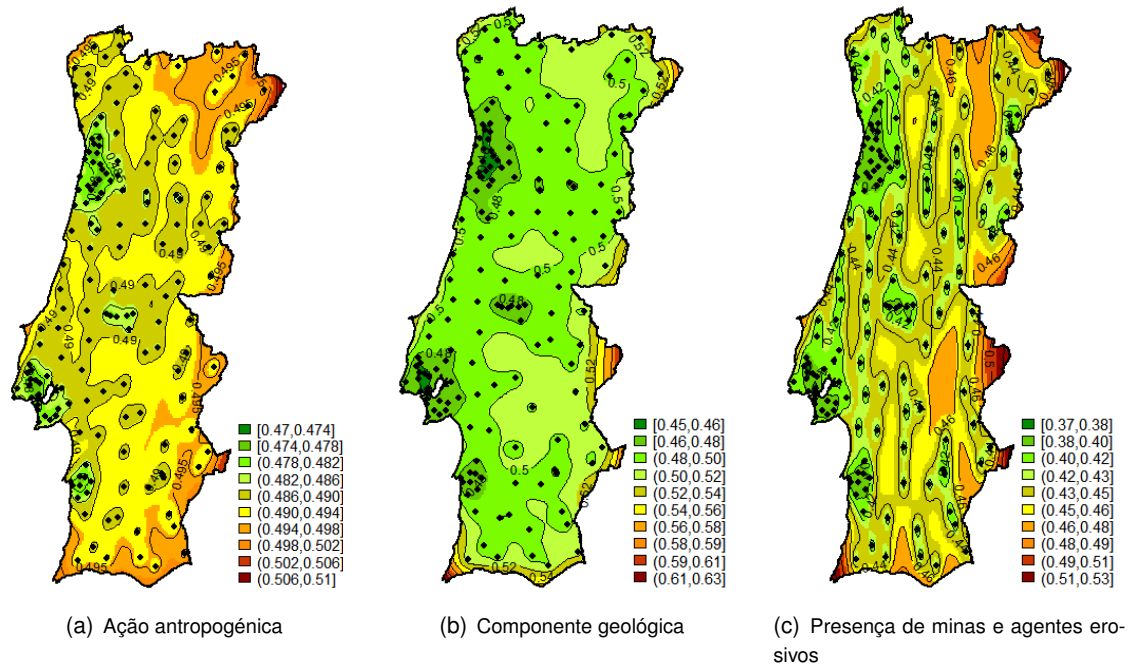


Figura 47: Desvios padrão associados a cada valor predito das fontes de poluição subjacentes à campanha de 1992.

Campanha de 1997

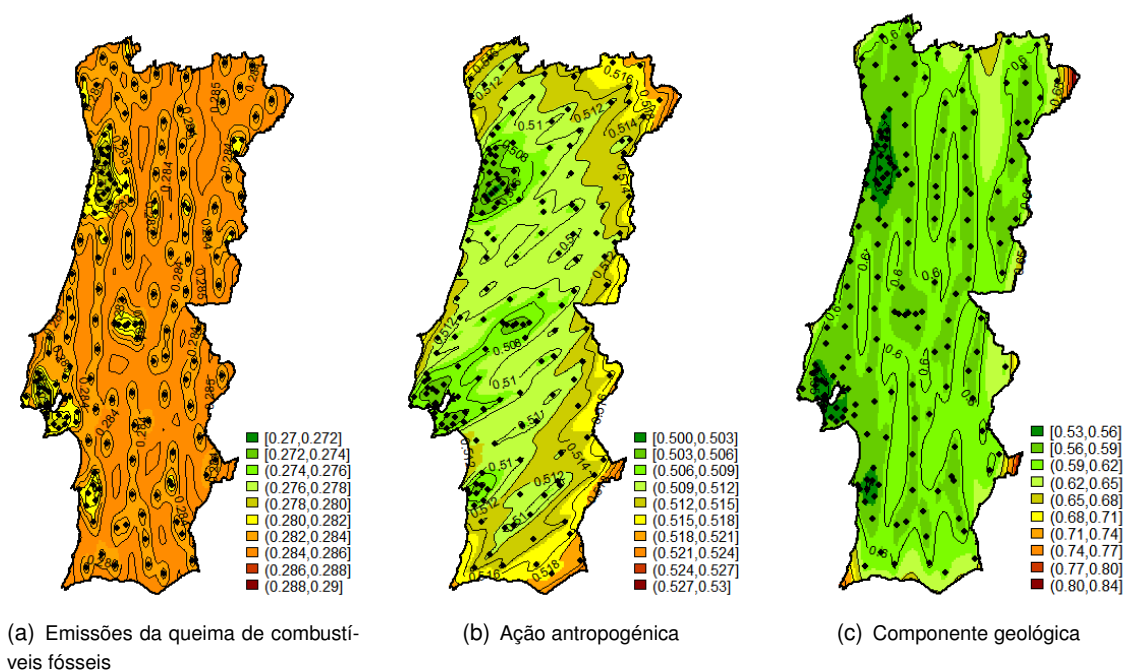


Figura 48: Desvios padrão associados a cada valor predito das fontes de poluição subjacentes à campanha de 1997.

Campanha de 2002

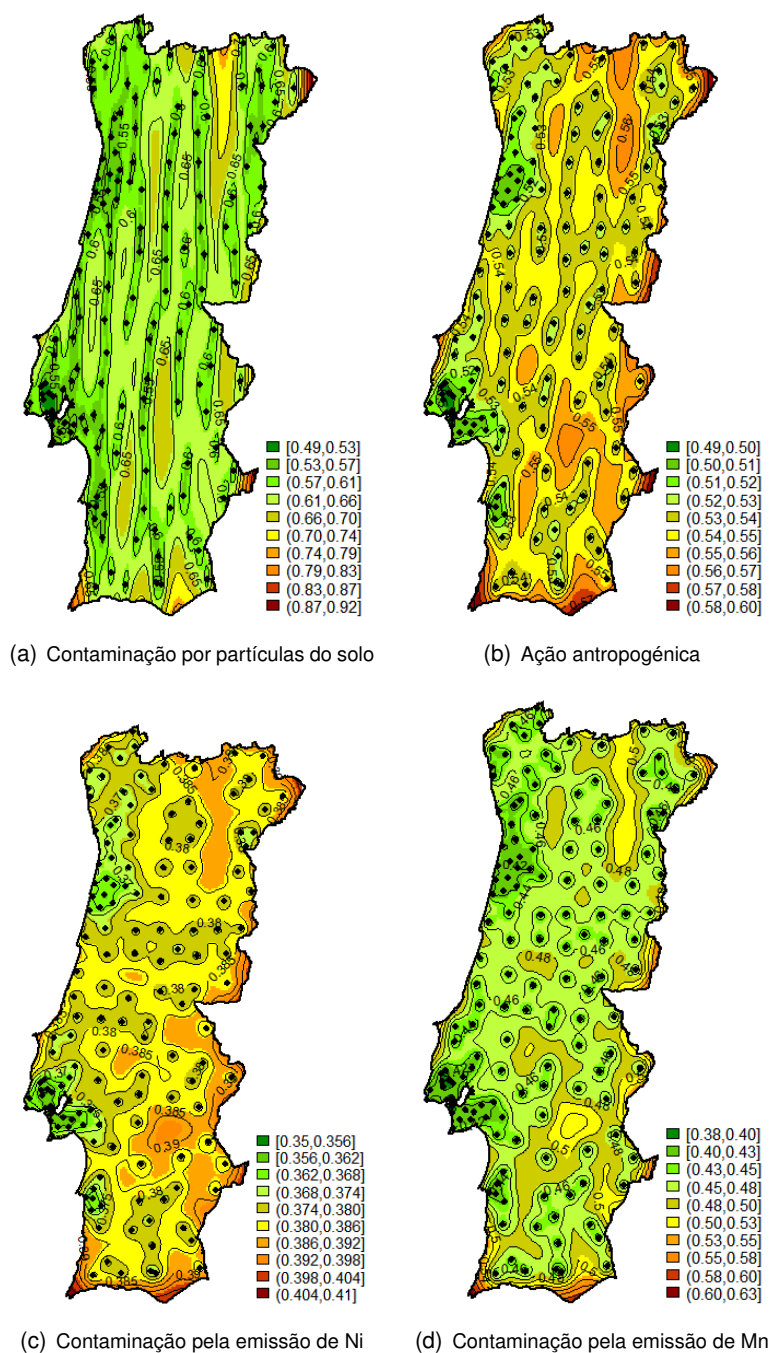


Figura 49: Desvios padrão associados a cada valor predito das fontes de poluição subjacentes à campanha de 2002.